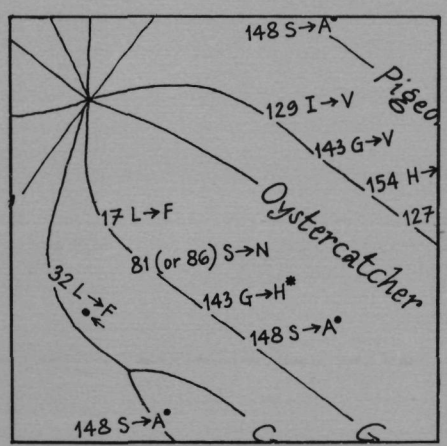
Crystallins and Avian Phylogeny



S. Stapel

Crystallins and Avian Phylogeny

This work was carried out under the auspices of the Foundation for Fundamental Biological Research (BION), with financial aid from the Netherlands Organization for Pure Research (ZWO) in the Department of Biochemistry (Head: Prof. Dr. H. Bloemendal), Faculty of Science, University of Nijmegen, The Netherlands.

Crystallins and Avian Phylogeny

Proefschrift

ter verkrijging van de graad van doctor
in de wiskunde en natuurwetenschappen
aan de Katholieke Universiteit te Nijmegen,
op gezag van de Rector Magnificus
Prof. Dr. J.H.G.I. Giesbers,
volgens besluit van het College van Dekanen
in het openbaar te verdedigen
op donderdag 6 maart 1986
des namiddags te 4 uur

door

Steven Olivier Stapel
geboren te Amsterdam

Promotor:	Prof. Dr. H. Bloemendal
co-referenten:	Dr. W.W. de Jong
	Dr. J. Wattel

. . . . . de heele wereld
om ons heen blonk in 't licht, de aarde gaf licht op
en zoo ver als onze oogen reikten was de wereld van
ons, en verder.

NESCIO, "BUITEN-IJ"

aan: M.C. van Riemsdijk

voor: Saskia

DANKWOORD

"Een proefschrift schrijf je niet alleen", stelde een zeer gewaardeerd ex-collega in 1983. Misschien schrijf je het wel alleen, maar zonder de betrokkenheid van anderen valt er meestal weinig van te verwachten.

Daarom dank ik Anneke Zweers voor haar introducties in de technische kanten van het onderzoek en de prettige samenwerking, en Marlies Versteeg voor de vele aminozuur analyses en haar assistentie bij het "Edmannen". Ook Gerrit Groenewoud moet vermeld worden voor zijn eigen specifieke bijdragen, vooral op kampeergebied.

De bijdragen van Huub Dodemont, Jaap Kan (Leiden) en Jack Leunissen krikten het wetenschappelijk gehalte enigszins op, waarvoor ik hen zeer erkentelijk ben.

Natuurlijk dank ik ook mijn collega's van het eerste uur, Guy Berbers, Theo Cuypers en Huub Driessen, die nooit moe leken te worden om mij op eiwit-chemisch gebied voor te lichten en met wie ik een uiterst gezellige tijd heb beleefd.

Ook de contacten met de "Hoenders-boys" en de "Lunch-club" droegen hier sterk aan bij.

De uitstekende service mij verleend door een aantal medewerkers van het "Dierenlab" en de afdeling Medische Fotografie/Illustratie mag ook beslist niet onvermeld blijven.

Uiteindelijk is het de straffe hand van Els van Genne die ervoor zorgde dat er nu iets te lezen valt.

Gerrit Stapel en Theo Peeters droegen bij aan de visuele aspecten van dit boekje, terwijl Susan Metselaar een taalkundige bijdrage leverde.

Zonder de uitgebreide steun van een aantal emotioneel betrokkenen had het nooit iets kunnen worden. Hiervoor wil ik in het bijzonder Anke, Maud en Saskia bedanken.

CONTENTS

PREFACE

"He who wants to ride two camels at the same time will end up in the desert sand", is an old Arabian saying which perfectly reflects the troubles occurring to someone who has to operate within the tangent plane of two different scientific disciplines.

This may easily be exemplified by the work described in this thesis: taxonomic questions are to be answered by biochemical approaches. This will undoubtedly lead the taxonomist to find certain insufficiences from a taxonomic point of view, while the biochemist will wrinkle his eyebrows by the, in his opinion, excessive taxonomic discourses.

Nevertheless, one should always keep an open eye to the possible solutions offered by other scientific disciplines, since the attempts of a "non-prejudiced outsider" from another scientific field may sometimes lead to long wished-for answers. I therefore hope that the results of this work will lend support to a positive appreciation of interdisciplinary scientific approaches, and will also stimulate the interest in further studies of avian phylogeny by comparative macromolecular sequence analyses.

SUMMARY

Determination of avian phylogenetic relationships on the basis of morphological and anatomical characters has appeared to be, in a number of cases, very troublesome. This has to be attributed to the circumstance that a proper taxonomic assessment of these characters may be utterly difficult (Chapter I).

In order to provide some additional taxonomic evidence, it was tried to find biochemical characters which might be useful for the clustering of avian species.

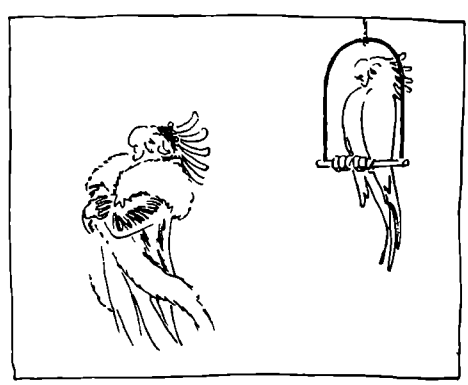
The discovery of two novel lens proteins, ε-crystallin and 48 KDa-crystallin (Chapter II), both appearing scattered among the 21 investigated avian species, initially seemed to provide promising characters for the grouping of species, but it turned out to be impossible to perform cladistic analyses, using these proteins as shared derived characters, since both components were also found in the reptilian "outgroup". It was obvious that phenetic approaches would not lead to phylogenetically relevant constructions.

Taxonomically important conclusions, on the other hand, could be drawn from comparative α-crystallin A amino acid sequence determinations, involving 21 avian species, which strongly indicated a monophyletic origin and sister group relationship to the other birds for the ratites, the large flightless birds of the Southern Hemisphere (Chapter III).

A great number of avian phylogenetic problems, however, cannot be solved by αA-analyses, due to its slow evolution rate. We therefore tested the suitability of another eye lens protein, δ-crystallin, and found it to be a promising macromolecular tool for the solution of obscure avian interordinal relationships (Chapter IV).

Chapter 1

Introductory Avian Biology



I-1: Biological classification

For many centuries man has tried to gather knowledge about the hierarchical structure of nature, and up till now one of the most appealing challenges that has remained is to achieve a complete classification of biological organisms.

Such a classification, which will often markedly differ from groupings based on relatively superficial characters like colour, size or habitat, now tends to be considered of maximum value when evolutionary relationships are being expressed. This is the main reason why an increasing number of biologists has taken renewed interest in the study of higher taxonomic categories, in order to reconstruct their evolutionary history and genealogical relationships.

The classification of biological species can, in many cases, be relatively easily performed on the basis of external appearances. Starting from the assumption that those biological manifestations which show the greatest "overall similarity" are most closely related, it turns out not only to be possible to assign, for instance, some feathered flying animal to the class Aves, but also to indicate if it concerns some duck-like, stork-like or hummingbird-like representative of this class.

It is by this almost obvious, intuitive way that biological species can be positioned together in a hierarchical taxonomic system.

Determination of the phylogenetic relationships of such groups, however, can sometimes be very troublesome because more distantly related taxa offer fewer possibilities for direct comparison of external characters. Moreover, it is never possible to obtain positive experimental proof for the alleged evolutionary reconstructions.

To provide a probable explanation for the overwhelming variety of biological species we know today, evolutionary biologists have to trace back the evolutionary events leading to this biological diversity, and they can essentially make use of two different approaches for the clustering of biological species (see: Mayr, 1981).

1) By methods according to the principles of "phenetic taxonomy", which group biological species on the basis of "overall similarity". By applying this approach, as many characters as possible are compared, followed by extensive judgement of their degree of resemblance. Although the preliminary grouping of taxa is nearly always based on phenetic evaluation of similarity, this approach does not provide a satisfactory test to a taxonomic theory in critically difficult cases. It is often very difficult to determine the degree of similarity required for the proper clustering of different taxa, or to decide which taxonomic importance must be attributed to similar traits.

2) Another possible approach to turn to, may be the application of cladistic analysis.
According to this method, taxa are positioned together because they share uniquely evolved, i.e. derived characters, or "synapomorphies" (Hennig, 1966), which set them apart from all other taxa. These characters are therefore considered to be of great taxonomic weight. This method, based upon the principle that "... the joint possession of homologous derived characters proves the common ancestry of a given set of species" (Mayr, 1981), may lead to the construction of a continuously bifurcating branching diagram, reflecting the patterns of phylogenetic relationship.

Both approaches are characterized by specific pitfalls, the most important one being a possible erroneous clustering on the basis of apparently similar characters, generated by convergent evolution. This is the reason why it will always be one of the most important tasks of the systematic biologist to recognize "homologies", and to distinguish the characters which are of taxonomic value.

Although in most cases both approaches will result in a similar taxonomic classification, it must be stressed that application of both methods to the very same problem may sometimes lead to different classifications, as can be seen in Figure 1.

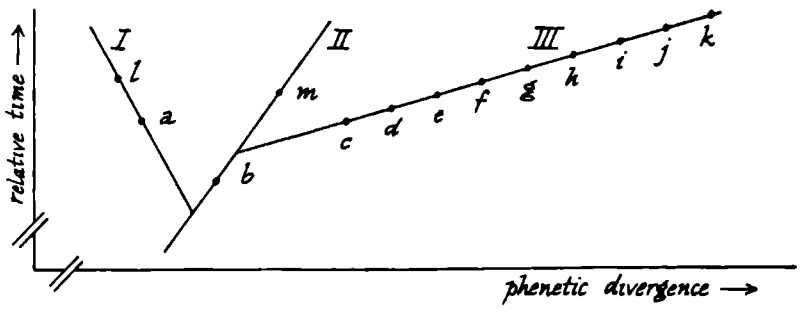


Figure 1: Cladogram of taxa I, II and III. Essentially after Mayr (1981).

The cladogram combines taxa II and III because of the circumstance that they both possess shared derived character b, which is not found in taxon I.

According to phenetic principles, taxon III is far removed from taxa I and II, on the basis of the discriminating characters c to k, which are only found in taxon III.

These unique characters, designated as apomorphies (Hennig, 1966) determine the peculiarity of a given taxonomic group. This can be exemplified by the avian assemblage, which has, after splitting off from the reptiles, developed a number of unique characters, definitely discriminating the birds from their close reptilian relatives. Although on the basis of overall similarity (i.e. on phenetic grounds), birds are considered to be the "sister group" of all reptiles, cladistic analyses, based upon unique similarities exclusively shared by birds and crocodiles, clearly point to a sister group relationship for these latter two groups, placing them away from the other reptiles (Walker, 1972; Whetstone and Martin, 1981). Since useful information can be obtained from cladistic analyses, based upon a limited number of characters, this approach might be very suitable to test the outcome of phenetic analyses, and in some cases unravel problems which could not be solved by the other method. Phenetic analysis, on the other hand, seems to be helpful as many characters are involved. However, testing of a cladistic classification by phenetic methods is logically unsound.

One of the most difficult tasks of the taxonomic biologist performing cladistic analyses, however, is the distinction of the required shared derived characters, since convergent evolution may seriously trouble the procedure.

Therefore, the essential requirement for proper cladistic analysis is the outgroup comparison. If some character, found in some taxa within a given taxonomic group is used for the grouping of these taxa, it is very important to be sure that this specific trait is not present in taxa outside of the group, which would designate this character to be primitive, or plesiomorph (Hennig, 1966), rendering it virtually useless for the purpose of delineating the evolutionary history of a certain group.

Throughout the many years of study, the reconstruction of the evolutionary history of biological species has caused the biologist a great deal of trouble. This can be seen from the circumstance that well-established opinions may be thrown over when new evidence seems to point to different phylogenetic conclusions, resulting in a great number of unsettled taxonomic problems, attended with vigorous discussions and bitter disagreements.

All biological classes are characterized by some major taxonomic problems, which are difficult to solve since the available morphological evidence does not lead to unanimous interpretation.

In this treatise some major taxonomic problems of the class Aves, the living birds of the world, will be outlined, as well as the usefulness of biochemical approaches for their solution.

I-2: Avian origin and taxonomy

Although an impressive amount of information is available about the distribution, habitat requirements, life cycles etc. of the approximately 8600 avian species recognized today, a well-known ornithologist stated in 1971:

> "The living birds are the best-known group of animals, but their origin, history and phylogeny are very poorly documented in comparison with the other vertebrates. More than a century has passed since the discovery of the first skeleton of Archaeopteryx, but still unknown are the links connecting this momentous find to its reptilian ancestors on the one hand, and to its avian descendants to the other" (Brodkorb, 1971).

This lack of knowledge has been attributed by another avian biologist to three main reasons:

1. The restrictive physical demands of flight, causing mass convergence,
2. The shortage of morphological features (such as teeth) that vary sufficiently among the major groups to be of useful taxonomic value, and
3. The lack of a good fossil record, because of the fragility of the bones and the arboreal nature of most avian species (Feduccia, 1977).

Although Brodkorb (1971) stated that the lack of fossil evidence may have to be attributed to the small number of paleornithologists studying the fossil record, it is generally agreed that the major bottleneck preventing the proper reconstruction of avian evolutionary history is found at the first point, indicating the importance of convergent evolution. This has led the dean of paleobiology, A.S. Romer, to the following statement (Romer, 1966):

> "... birds of today despite their varied plumage, songs, and habits, are very similar to one another in their structure. They are divided into many orders; but the differences, for example, between a hummingbird and an albatross are much less than those between a seal and cat ... The different bird orders have, in general, no more differences between them than exist between families in other classes of vertebrates, and anatomical differences between bird genera are often so slight that fossils are hard to place."

Statements like these have influenced avian systematists to a great extent, which finally seems to have resulted in some kind of stalemate position concerning two major problems: the assumed age of the avian assemblage and the phylogenetic interrelationships of the avian orders (Table 1). Although arguments and discussions among ornithologists have continued for a long time, it has not been possible so far to reach unanimous opinions about these two problems.

To illustrate this, some of the current opinons on the matter will be outlined below.

A. The origin of birds

One major problem concerning the avian assemblage has always remained the question of the origin of birds, the answer of this matter being directly connected with the question of the age of the avian stock.

After more than a century of studying the fossil evidence, no unanimity has been reached among avian biologists, but two main theories, conflicting with respect to the specific lines of descent and terms of the time when the first bird appeared, are nowadays both supported by various taxonomists.

Since in both views birds are assumed to be derived from reptilian ancestors, although in completely different ways, a short overview (couched in "evolutionary", not cladistic terms) of the assumed reptilian evolution is given below (Romer, 1966).

After the separation of the paleozoic reptiles in the Pelycosaurs (a branch which led to the mammalian stock) and the Cotylosaurs, the latter group gave rise to, on one hand, the turtles (Chelonia), lizards and snakes (Lepidosauria), and on the other hand the Permian thecodont reptiles. Among these latter organisms we find the so-called Pseudosuchians (230 million years ago), from which group, according to one of the stem theories, the pterosaurs (non-avian flying reptiles), the crocodiles and dinosaurs are thought to be derived.

The dinosaurs then are believed to have split into the so-called Ornithischia and Saurischia, of which the latter group gave rise, in the early Jurassic, to the herbivorous Sauropoda (with a.o. Brontosaurus) and the carnivorous Theropoda. The theropods split into the Carnosauria (with Tyrannosaurus) and the smaller Coelurosauria (with Compsognathus and Struthiomimus).

According to the first of the two main theories on avian descent, originally put forward by Huxley (1868), the first known birds showed a striking resemblance to the coelurosaurian reptiles, and for this reason the oldest bird, Archaeopteryx, was initially considered to be such a reptile. For this reason, one of the two theories about avian ancestry is generally referred to as the "coelurosaurian-ancestor theory".

Although many more theories emerged (see: Martin, 1983), there is one other main theory which had gained a lot of support for many years.

It was as early as 1877 that the dinosaurs were considered too diverse and too specialized to be the progenitors of the birds (Marsh, 1877). This would imply an independent origin for birds and dinosaurs, and in 1913 Broom postulated his "pseudosuchian-ancestor theory" on the basis of the early Triassic fossil Euparkeria, found in South Africa, which was considered to represent an early stage of the archosaur radiation, sufficiently primitive to be the progenitor of both dinosaurs and birds. This theory was propagated by Heilmann in his influential book "The Origin of Birds", in 1926, and supplanted for some decades the "coelurosaurian ancestor theory", until in the 1970's the smouldering controversies rose again through recent papers by Ostrom (1976). His conclusion, after intensive study of the five known skeletal specimens of Archaeopteryx, was that avian evolution has run along the following course: Pseudosuchia → Coelurosauria → Archaeopteryx → higher birds, which designates the living birds to be living dinosaurian descendants.

Papers by Galton (1970), suggesting a close ornithischian-avian relationship; Walker (1972), bringing up the relatively specialized pseudosuchian Sphenosuchus as common avian-crocodilian progenitor and Gardiner (1982), with his assumed avian-mammalian sister group relationship, complicated the picture

of avian ancestry even further.

Apart from the question of designating the true avian progenitor, the two main theories on the origin of birds, as outlined in figure 2, bear some radical implications concerning the problem of the age of the avian stock. Acceptance of the "coelurosaurian-ancestor theory" means that the ancestral bird must have originated as a separate evolutionary lineage about 140 million years ago, while the "pseudosuchian-ancestor theory" assumes this oldest avian ancestor to have come into existence about 230 million year ago.

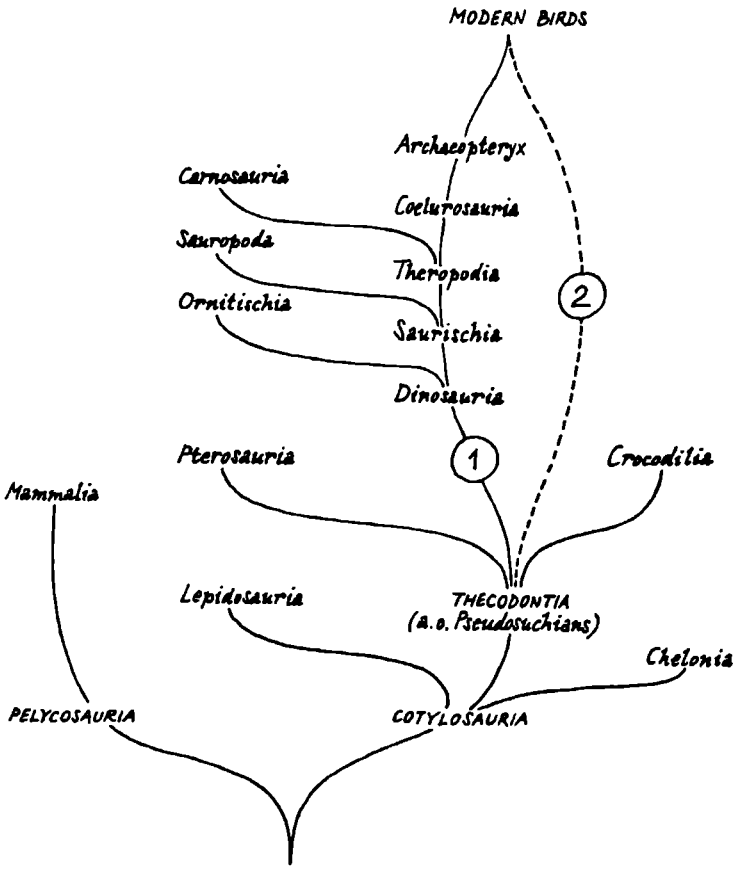


Figure 2: Generally accepted theories about avian origin.
1. "Coelurosaurian-ancestor theory"
2. "Pseudosuchian-ancestor theory"

B. Avian phylogeny

The uncertainties about avian ancestry may be mainly attributed to the lack of sufficient conclusive fossil evidence; this is not true for the problems of the avian interrelationships.

Birds can be divided into a number of orders without too many difficulties (table 1), but the conclusive determination of the phylogenetic relations of these orders has turned out to be an almost impossible task, taking into account the efforts put into it by a great number of ornithologists. Although all the required anatomical-morphological evidence is at hand, it seems to be impossible, in spite of many years of argument, to establish some general unanimity about some of these interordinal problems.

A number of these questions are outlined below, providing some demonstration of the deep disagreements found among ornithologists.

The Ratites

A very special assemblage among the living birds is the group of the ratites, the large flightless birds of the Southern Hemisphere, comprising ten living species. They include the ostrich (genus Struthio) of Africa and Arabia; the two species of South American rheas (genus Rhea); the Australian emu (genus Dromaius); the three Australo-Papuan cassowaries (genus Casuarius); and the three kiwi-species of New Zealand (genus Apteryx). Their phylogenetic relationships to other birds and especially to the South American tinamous have been the subject of numerous publications.

Another major problem has troubled the ornithologist for more than a century: do the ratites (including the 47 extinct species known today) represent a monophyletic assemblage or did they emerge independently from different ancestral flying progenitors?

A review of the literature concerning this problem by Sibley and Ahlquist (1981) clearly shows the lack of concensus throughout the years:
Huxley (1867), on the basis of the structure of their bony palate, considered them to be monophyletic; Führbringer (1888), on the contrary, regarded this palatinal structure as the result of convergent evolution, concluding the ratites to have originated independently; Pycraft (1900), although considering the ratites to be polyphyletic, decided to place them together in one group, while Stresemann (1934) and Wetmore (1930), whose classifications have been generally accepted for the last decades, followed Führbringer in his idea of independent ratite ancestry; Bock (1963), on the basis of cranial anatomical evidence positioned them together again. The same proposition was made by Parkes and Clark (1966) for reasons of unique similarities in their rhamphothecal structure, while Cracraft (1974) also supports their monophyly on the basis of hind limb and pelvic structures. Yet, Storer (1971) wrote:

> "On zoogeographic grounds, ratite birds must have originated independently from flying ancestors at least four times."

and:

> "... it is unwise in a phylogenetic classification to place them in one separate group",

once again underlining the profound differences of opinion about this question.

Table 1: Avian Orders

Biological species are classified in a hierarchical structure which is characterized by the following taxonomic levels: species, genus, family, order and class. However, intermediate levels, like subgenus and superfamily, can also be inserted.
The number of orders into which birds are devided has varied widely, as outlined by Sibley and Ahlquist (1981), but current classifications comprise approximately 25 avian orders. The scientific and common names of 25 orders and some of their representatives are given below, according to Peters et al. (1979).

Order	Common names of representatives	Dutch common names
Struthioniformes	ostriches, rheas emus, cassowaries	struisvogels, nandoes, emoes, casuarissen
Tinamiformes	tinamous	tinamous
Procellariiformes	albatrosses, petrels	albatrossen, stormvogels
Sphenisciformes	penguins	pinguins
Gaviiformes	loons	duikers
Podicipediformes	grebes	futen
Pelecaniformes	pelicans, cormorants, gannets, frigate birds	pelikanen, aalscholvers, jan-van-genten, fregatvogels
Ciconiiformes	herons, storks, ibises	reigers, ooievaars, ibissen
Phoenicopteriformes	flamingos	flamingos
Falconiformes	birds of prey	roofvogels
Anseriformes	ducks, geese, swans	eenden, ganzen, zwanen
Galliformes	chickens, turkeys, pheasants	hoenders, kalkoenen, fazanten
Gruiformes	cranes, rails	kraanvogels, rallen
Charadriiformes	waders, gulls, auks	steltlopers, meeuwen, alken
Columbiformes	pigeons	duiven
Psittaciformes	parrots	papegaaien
Cuculiformes	cuckoos	koekoeken
Strigiformes	owls	uilen
Caprimulgiformes	nightjars	nachtzwaluwen
Apodiformes	swifts, hummingbirds	gierzwaluwen, kolibris
Coliiformes	mousebirds	muisvogels
Trogoniformes	trogons	trogons
Coraciiformes	rollers, kingfishers	scharrelaars, ijsvogels
Piciformes	woodpeckers	spechten
Passeriformes	swallows, thrushes, crows, finches, etc.	zwaluwen, lijsters, kraaien, vinken, etc.

Additional data are provided by Sibley (1960), who concluded a common ancestry for the ostrich, cassowary and emu on the basis of egg white protein electrophoretic patterns; by Prager and Wilson (1976) on the basis of comparative immunology of a number of avian proteins and by De Boer (1980), concluding a monophyletic origin by karyotypical comparisons.

It turns out that a lot of evidence speaks for a ratite monophyly, but the phylogenetic position of this group in relation to the other avian orders remains unclear. This may be attributed to the circumstance that the comparative anatomical-morphological approach does not seem to result in unanimously acceptable traits linking the ratites to another avian group, while the applied molecular approaches by themselves do not provide valuable character states which reveal the identity of the sister group of the ratites among the living birds. This matter will be commented on later.

The ancestry and taxonomic position of the Anseriformes

The order Anseriformes, or duck-like birds and relatives, are traditionally considered to comprise two different families: the Anatidae, with the ducks, swans and geese; and the South American Anhimidae ("screamers"), placed in this order mainly by lack of a better alternative (Sibley and Ahlquist, 1972).

The anseriform ancestry has always remained obscure, but some tendency has existed to place the order close to the Ciconiiformes (herons and storks) (see: Sibley and Ahlquist, 1972), although the Galliformes (chicken-like birds) have also been suggested to be closely related to the Anseriformes (for an extensive literature survey: see Olson and Feduccia, 1980). Some authors then position the Anhimidae as an intermediate group between the Anseriformes and Galliformes (see: Bock, 1970). Olson and Feduccia strongly opposed a close anseriform-galliform relationship, underlining their rejection of this assumption with the statement that:

> "... Anseriformes differ from Galliformes in almost every anatomical feature and there is not the slightest resemblance between the two groups in their postcranial osteology."

This seemed to leave the question unanswered again, but fortunately, the problem was claimed to be settled by the finding of the Eocene fossil Presbyornis (Olson and Feduccia, 1980). This organism, "combining the body of a shorebird with a duck-like head" appears to be a true missing link without which the origins of the Anseriformes might have remained obscure indefinitely". It is considered by the authors to be a product of mosaic evolution, and because of its charadriiform characters they state that "the Anseriformes have descended from charadriiform ancestors".

This conclusion could never have been drawn without the finding of this particular fossil, since "... the Anseriformes, as far as known, differ so much from their charadriiform ancestors in their postcranial anatomy as to have disguised their origins up to the present".

Although it may be possible that the secrets of anseriform ancestry have now been revealed (or could it be equally possible to assume a charadriiform origin from anseriform representatives?), the intriguing points remain, that, as the authors state, neither the Galliformes nor the Charadriiformes show any resemblance to the Anseriformes in their postcranial anatomy, and that up till now nobody has ever found any morphological-anatomical reasons to assume a special relationship between Anseriformes and Charadriiformes.

It therefore does not seem reasonable to reject the possibility of a galliform-anseriform relationship for the very same reason, and because of the circumstance that

"... we cannot imagine any reasonable hypothetical sequence of evolutionary events that could produce a duck from a galliform" (Olson and Feduccia, 1980).

One of the questions that arise is, whether the authors would ever have imagined a reasonable hypothetical evolutionary sequence leading from a charadriiform to a duck, if Presbyornis had not popped up.

Cracraft (1981), who did not believe in the suggested mosaic evolution proposed by Olson and Feduccia, was not really able to prove the improbabilities of their assumptions. However, an evolutionary sequence leading to characteristic cranial structures in some ancestral branch without anatomical consequences for its postcranial morphology seems to be a remarkable phenomenon. It has to be admitted that the finding of more of these "mosaics" could be of great help to the solution of phylogenetic problems, but preferably no anseriform-galliform mosaics, or charadriiforms with chicken-like heads.

Similar problems as have been outlined for the ratites and the anseriformes can be spotted in many other avian orders. Many examples of such problem cases and conflicting opinions can be found in Sibley and Ahlquist (1972), and Cracraft (1980).

Another much discussed problem is, for instance, the alleged monophyly of the loons (Gaviiformes) and grebes (Podicipediformes). As has been reviewed by Cracraft (1982), there was initially a general consensus about the monophyly of loons, grebes and the extinct Cretaceous Hesperornithiformes. Around the beginning of the century, however, the opinions changed. Storer (1971), propagated a polyphyletic origin for the foot-propelled divers, while Cracraft (1982) positioned loons and grebes, together with the Sphenisciformes and Hesperornithiformes, in one of his taxonomic "divisions".

This led Olson (1982) to state:

"Not a single synapomorphy is advanced to justify "Division 1" as a monophyletic group."

That very same year Cracraft claimed to attribute eight synapomorphies to the loons and grebes (Cracraft, 1982), which might shift the general opinion again to a monophyletic assemblage of gaviiform and podicipediform birds.

What will be the final taxonomic position of the New World vultures (Cathartidae)? They are now traditionally placed among the birds of prey (Falconiformes). Nevertheless, their position has been challenged for more than a century (Konig, 1982). Ligon (1967) placed them on the basis of morphological characters close to the Ciconiiformes (herons and storks). Wolters (1975) positions them, in their own order Cathartiformes, between the Sagittariiformes (Secretary birds) and Ciconiiformes. Konig (1982), on the basis of ethological arguments ("Schnabeln") also places them close to the Ciconiiformes. It is to be expected that, as more evidence seems to point to a ciconiiform relationship, the New World vultures will not be placed with the birds of prey, and end as a separate order. The debate concerning this matter is still going on.

C. Conclusion

As can be concluded from the foregoing, it seems to be very difficult to reach a general consensus about a number of taxonomic problems concerning the avian stock.

Literature survey points out that phenetic analyses, based upon the available fossil material and anatomical-morphological characters, lead in a number of cases to conflicting taxonomic conclusions.

When, for example, the widely accepted idea of the avian-reptilian sister group relationship is suddenly challenged to favour a closer avian-mammalian relation (Gardiner, 1982), a lot of disquiet may occur among comparative morphologists and paleontologists, resulting in renewed evaluations of evidence, already reevaluated many times.

It must be doubted, however, if it will ever be possible to determine the relation of the scarce fossil material to the living taxa, or to decide whether Archaeopteryx represents a real avian progenitor or a side-branch of the avian line.

In the light of the problems outlined above, it will be clear that it would be very convenient to obtain new clues for the elucidation of avian taxonomic relationships.

REFERENCES-I

- Bock, W.J. (1970) in: "XV Congressus Internationalis Ornithologicus Abstracts": 66-67.

- Brodkorb, P. (1971) in: "Avian Biology" (Farner, D.S. & King, J.R. eds.) Academic Press, New York and London.

- Broom, R. (1913) Proc. Zool. Soc. Lond. 1913: 619-633.

- Cracraft, J. (1981) Auk 98: 681-714.

- Cracraft, J. (1982) Syst. Zool. 31: 35-56.

- De Boer, L.E.M. (1980) Nature 287: 84-85.

- Feduccia, A. (1977) Syst. Zool. 26, No. 1: 19-31.

- Galton, P.M. (1970) Evolution 24: 448-462.

- Gardiner, B.G. (1982) Zool. J. Linn. Soc. 74: 207-232.

- Hennig, W. (1966) "Phylogenetic Systematics" (Davis, D.D. & Zangerl, R., eds.) Univ. of Illinois Press, Urbana.

- Heilman, G. (1926) "The Origin of Birds", London: Witherby.

- Huxley, T.H. (1868) Ann. Mag. Nat. Hist. Lond. 2: 66-75.

- Konig, C. (1982) J. Orn. 123: 259-267.

- Ligon, J.D. (1967) Occas. Pap. Mus. Zool. Univ. Mic. 651: 1-26.

- Marsh, O.C. (1877) Proc. Am. Assoc. Adv. Sci. 1877: 211-258.

- Martin, L.D. (1983) in: "Current Ornithology Vol. 1" (Johnston, R.F., ed.) Plenum Press, New York and London.

- Mayr, E. (1981) Science 214: 510-516.

- Olson, S.L. & Feduccia, A. (1980) Smiths. Contr. Zool. 323: 1-24.

- Olson, S.L. (1982) Auk 99: 733-739.

- Ostrom, J.H. (1976) Biol. J. Linn. Soc. 8: 91-182.

- Peters, J.L. et al. (1979) "Check-List of Birds of the World" Vol. 1, 2nd edition (Mayr, E. & Cottrell, G.W., eds.) Museum of Comparative Zoology, Cambridge-Massachusetts.

- Prager, E.M. & Wilson, A.C. (1976) J. Molec. Evol. 9: 45-57.

- Romer, A.S. (1966) "Vertebrate Paleontology", Univ. of Chicago Press, Chicago, 3th ed., p. 468.

- Sibley, G.C. (1960) Ibis 102: 215-284.

- Sibley, G.C. & Ahlquist, J. (1972) Bull. Peabody Mus. Natl. Hist. 39: 1-276.

- Sibley, G.C. & Ahlquist, J. (1981) in: "Evolution Today" (Scudder, G.G.E. & Reveal, J.L., eds.) Proc. Sec. Int. Congr. Biol.: Hint Inst. Bot. Doc., Pittsburg, Pa.: 301-335.

- Storer, R.W. (1971) in: "Avian Biology" (Farner, D.S. & King, J.R., eds.) Academic Press, New York and London.

- Walker, A.D. (1972) Nature 237: 257-263.

- Whetstone, K.N. & Martin, L.D. (1979) Nature 279: 234-236.

- Wolters, H.E. (1976) "Die Vogelarten der Erde", Part 2, Paul Parey, Hamburg und Berlin.

Chapter II

Avian Phylogeny at the Molecular Level

II-1: Molecular studies of avian relationships

Avian phylogeny has not only been studied at the anatomical-morphological level. Other biological disciplines have been involved as well, such as studies of specific parasites, ecology and ethology.

In the recent decades, however, a new approach has been introduced: taxonomic study at the molecular level, comprising comparative protein electrophoresis, immunological studies, amino acid sequence determination, and more recently, DNA-DNA hybridization, DNA restriction analysis, and, as a rising possibility, nucleic acid sequence determination.

The first major comparative study of electrophoretic protein patterns was carried out by Gysels (1964), who subjected the soluble proteins from eye lens, skeletal-, heart- and stomach muscle of 233 avian species to comparative electrophoresis in an agar-system. In his own opinion, the patterns he obtained with lens proteins did provide relevant information from a taxonomic point of view because of the circumstance that "... relatively little alterations did occur in the structure of lens proteins, and that perhaps less convergences than in other organs trouble our vision ...".

Together with the protein patterns the lens glycogen content was determined, leading to a grouping of specific patterns which was also assumed to represent taxonomically important information (see: fig. 3).

Although it was difficult to draw definite taxonomic conclusions, the author considered the discovery of a "Typical Songbird Component" of major interest. This component, however, a lenticular protein, was also said to be present in lenses of representatives of ciconiiform, falconiform, gruiform, strigiform, and coraciiform birds. Since Gysels, on the basis of his electrophoretic data, attributed a sphenisciform lens type to the guillemot Uria and a charadriiform lens type to the spoon-billed heron (Cochlearius cochlearius) etc., it is no wonder that the construction of a phylogenetic tree, expressing evolutionary relationships of different avian orders, could not be achieved on the basis of these data.

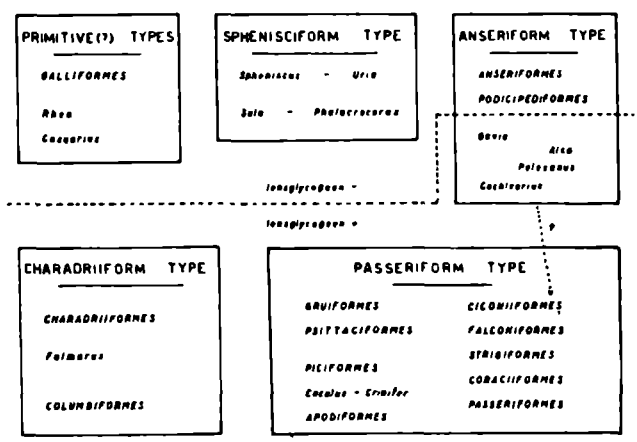


Figure 3: Lens patterns and inferred relationships by comparative lenticular protein agar-electrophoresis. After Gysels (1964).

Another example of this type of investigation was presented by Kitto and Wilson (1966), who compared the electrophoretic mobilities of a soluble enzymatic compound from the heart muscle, designated as S malate-dehydrogenase, of more than 100 avian species, in a starch-gel electrophoretic system. Their main conclusion was that the shorebirds (Charadriiformes), as well as the swifts and hummingbirds (Apodiformes), are homogeneous groups, which had been doubted by a number of ornithologists. For the rest, no phylogenetic inferences could be obtained from their data.

Other major comparative protein studies were undertaken by Sibley (1970), and by Sibley and Ahlquist (1972). After comparison of electrophoretic patterns of egg-white proteins of 816 species of birds, one of their conclusions was:

> "... the egg white patterns of the large ratites are more similar to one another than to those of any other avian group. From all available evidence it seems likely that these birds were derived from a common ancestor".

In a rather tentative way a close tinamous-galliformes association is suggested, and a number of possible interordinal relationships are indicated, without any further outspoken conclusions.

Although the comparative protein electrophoresis investigations outlined above formed the basis for some new taxonomic considerations, the validity of these techniques was not unchallenged, because of some serious drawbacks.

The first problem is, that the electrophoretic mobility of a given protein is dependent on the net charge of the molecule, and therefore on the number of acidic and basic amino acids which will be detected by the electrophoretic techniques, provided that no compensating opposite-charge substitutions take place. The equally informative charge-neutral substitutions escape observation.

Another difficulty with comparative protein pattern interpretation can be, that in some cases several charge-variants of the very same protein may originate, for instance by deamidation or degradation processes, which may further complicate the protein pattern.

For these reasons, the electrophoretic approaches described above do not seem to be very useful for the solution of taxonomic problems, and do not lead to informative phylogenetic tree constructions, since apomorph or plesiomorph characters cannot be distinguished.

Immunological data are claimed to be valid for the construction of a tree, expressing genealogical relationships (Prager and Wilson, 1976). These authors calculated immunological distances between a number of homologous proteins from different avian species by the micro-complement fixation technique. Their experiments involved 24 avian species, of which they compared the immunological distances of transferrin, albumin, ovalbumin and lysozyme. Some of their results are shown in figure 4.

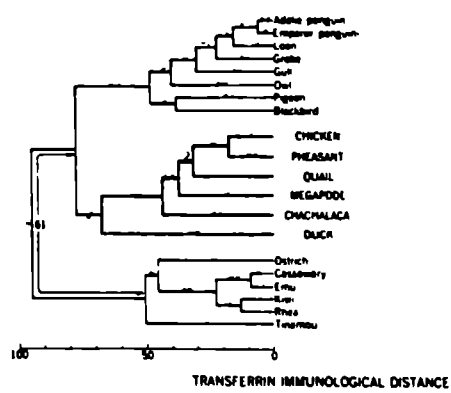


Figure 4: Avian phylogeny based upon immunological distances of avian transferrin.
From: Prager and Wilson (1976).

Especially notable are the calculated positions of the ratites, kiwi and tinamous, which are clustered together, and of the anseriform and galliform birds, in relation to the other birds.

Comparable results for the ratites, kiwi and tinamous were obtained by Sibley and Ahlquist (1981), on the basis of DNA-DNA hybridization experiments, as is shown in figure 5.

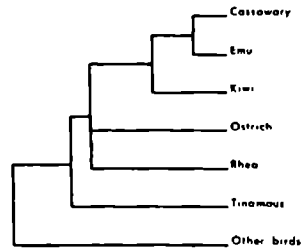


Figure 5: Phylogeny of the ratites, kiwi and tinamous.
From: Sibley and Ahlquist (1981). No specification of "other birds" is given in the original article.

This latter technique was also used by the same authors to shed some light on the obscure phylogenetic relationships of a number of songbirds (Sibley and Ahlquist, 1980).

The immunological distance determinations and DNA-DNA hybridizations result essentially in phenetic taxonomic information, and therefore cannot be used for the construction of cladograms. The constructed trees of figures 4 and 5 should therefore be designated as phenetic trees. The electrophoretic protein mobility comparisons (Gysels, 1964; Kitto and Wilson, 1966; Sibley and Ahlquist, 1972), which might have led to the recognition of shared derived characters, have also not provided the expected taxonomic breakthrough, and the conclusions about avian phylogenetic problems were vague and unsatifactory.

Consequently, an attempt will be made to perform a cladistic analysis in order to reconstruct the phylogeny of some avian orders, again using the presence of specific lenticular protein compounds as character states.

II-2: Novel avian lenticular proteins

The electrophoretic patterns obtained with muscular and lenticular proteins (Gysels, 1964), as well as with egg white proteins (Sibley and Ahlquist, 1972), are mainly the result of two physical protein parameters: net charge and molecular mass, and are also influenced by properties of the matrix material, being agar or starch. Identification of specific polypeptide chains may be rather precarious in these types of electrophoretic system.

A considerable step forward in the field of comparative electrophoresis was made by the introduction of the Sodium Dodecyl Sulfate Poly-Acryl Amide Gel Electrophoresis (SDS PAGE) by Laemmli (1970). By application of this technique, polypeptide chains are separated on the basis of their molecular mass in a synthetic polymeric matrix with high resolving power, resulting in a relatively easily interpretable protein banding pattern. In combination with protein isoelectrofocusing, very informative two-dimensional protein patterns can be obtained, as can be seen in the following section.
(For more detailed information about SDS-PAGE and isoelectrofocusing, see: Weber et al., 1972; O'Farrell, 1975).

Application of these electrophoretic techniques revealed the presence of two hitherto undescribed lenticular proteins in some avian species, which were isolated and subjected to a thorough biochemical characterization, in order to determine their relationship to the other water-soluble lens proteins, and also to provide some new, well-defined tools for comparative studies of gene expression.

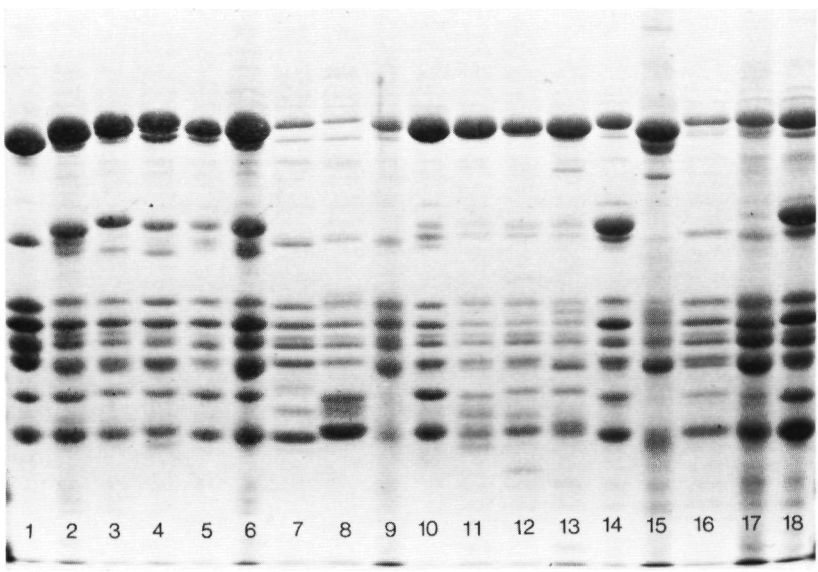


Figure 6: SDS-gel electrophoretic patterns of chicken (1), bean-goose (2), gadwall (3), common golden-eye (4), mute swan (5), tufted duck (6), woodpecker (7), cuckoo (8), coot (9), yellow billed loon (10), red-throated loon (11), common loon (12), arctic loon (13), heron (14), gentoo penguin (15), fulmar (16), curlew (17) and oystercatcher (18).

As can be seen in figure 6, characteristic polypeptide patterns can be obtained by application of this SDS-electrophoresis. Some anseriform patterns show distinct similarities, but differ markedly from, for instance, the patterns of woodpecker or cuckoo. Also the gaviiform pattern can easily be distinguished from penguin or oystercatcher.

Most striking, however, was the discovery of the two additional protein compounds "ε-crystallin" and "48 KDa-crystallin" by this technique, in some avian lenses. A physical-chemical characterization of these eye lens proteins is described in the following section.

Biochemical characterization of ϵ-crystallin and 48 KDa-crystallin

Eur. J. Biochem. 147, 129-136 (1985)

FEBS Lett. 162, 305-309 (1983)

Eur J Biochem 147 129–136 (1985)

ε-Crystallin, a novel avian and reptilian eye lens protein

Steven O STAPEL Anneke ZWEERS Huub J DODEMONT Jaap H KAN and Wilfried W de JONG

Laboratorium voor Biochemie Universiteit van Nijmegen

(Received July 2 1984) – EJB 84 0709

Gel filtration of Peking duck eye lens proteins reveals a component eluting just behind δ-crystallin and comprising approximately 10% of the total soluble protein The native M_r of this additional component is estimated to be 120000 it appears to be composed of three identical chains of M_r 38000 and pI 7 5 Circular dichroic spectroscopy showed a relatively high α-helical content No immunological cross-reactivity is found with α β- γ- or δ-crystallins and partial amino acid sequence determinations likewise failed to reveal any similarity with other known crystallins We conclude that this protein represents another and novel family of eye lens proteins for which we propose the designation ε-crystallin ε-Crystallin is translated from a 1450-base mRNA which has been partially purified ε-Crystallin is found scattered among avian and reptilian taxa but not in other vertebrates Its rate of evolutionary change seems to be as slow as that of α- and β-crystallins

The structural proteins of the vertebrate lens the crystallins have proven to be suitable objects for the study of a great variety of fundamental biological phenomena, like differentiation aging and evolution [1] They are also increasingly used as valuable tools to study gene structure, expression and regulation [2–4]

Three types of crystallins, α, β and γ, have classically been distinguished in virtually all investigated vertebrate lenses [5] The β- and γ-crystallins are now known to be structurally related [6] Bird and reptile lenses contain in addition δ-crystallin [2]

The impression might be gained that these are the only major lens-specific proteins in vertebrates However, in spite of the overall similarity in structure and function of the eye lens, great differences do occur in the relative proportions of the various crystallins while additional, in most cases poorly characterized components are present in several species [7] We recently described a monomeric 48-kDa crystallin occurring in lenses of lampreys some fishes reptiles and birds [8] Now we report the thorough characterization of another major lens protein which was initially observed in Peking duck, where it constitutes 10% of total lens protein It was subsequently identified, in varying amounts, in several other avian and reptilian species This protein designated as ε-crystallin, further extends the possibilities for molecular biological studies in the lens system

MATERIALS AND METHODS

Lenses were obtained from Peking duck (*Anas platyrhynchos*), tufted duck (*Aythya fuligula*), heron (*Ardea cinerea*) and American alligator (*Alligator mississippiensis*)

Abbreviations SDS PAGE sodium dodecyl sulphate/polyacrylamide gel electrophoresis CD, circular dichroism

Enzymes Thermolysin (EC 3 4 24 4) trypsin (EC 3 4 21 4)

Protein isolation and characterization

Lenses were isolated and treated as previously described [8] Gel filtration of lens extracts was carried out on a column (150 × 4 cm) of AcA 34 as described elsewhere [9] Slab gel electrophoresis and two-dimensional electrophoresis of water-soluble lens proteins was performed as described by Laemmli [10] and O Farrell [11] respectively

Immunological identification of lens components was performed with the immunoblotting procedure as described by Burnette [12] Rabbit antisera were obtained by immunization with ε-crystallin which, after gel filtration, had been further purified by preparative SDS-PAGE and was removed from the gel material by electroelution Rabbits were injected with 2 mg protein in the presence of complete Freund s adjuvant and blood samples were taken after 3 weeks The specificity of the antiserum was apparent from the fact that it reacted exclusively with ε-crystallin on two-dimensional immunoblotting of total duck lens protein (cf Fig 3C)

Native protein molecular mass determinations were obtained by subjecting 3 mg of peak fraction ε-protein and 6 mg of total lens extract to high-performance gel permeation chromatography on TSK GEL SW-type columns, in combination with detection by low-angle laser light scattering (LALLS) as described by Bindels et al [13] Subunit molecular masses were estimated from SDS-PAGE

Amino acid analyses and primary structure determination by applying the dansyl-Edman procedure to peptides obtained by tryptic digestion of cyanogen bromide fragments, were essentially as described by Driessen et al [6] CNBr fragments were fractionated by gel filtration over Sephadex G-50 sf in 5% acetic acid Fractions containing different CNBr fragments were digested with trypsin and resulting peptides separated by gel filtration (Sephadex G-50 sf in 0 1 M NH_4HCO_3) and or high-voltage electrophoresis at pH 6 5 followed in the second dimension by descending chromatography

Circular dichroic spectroscopy of millipore-filtered peak fractions of Peking duck crystallins, at a concentration of 1 mg protein/ml in 1% NH_4HCO_3, pH 7.9, was carried out on a CNRS Roussel-Jouan III dichographe (Jobin-Yvon, France), after instrument calibration with (+)-10-camphor sulfonic acid and D-pantoyl lactone at 20 °C and with a scan rate of 0.05 nm/s. CD spectra were digitised by hand every 2 nm in the 182–250-nm range, which data were used as input for a computer-programmed calculation (IBM-supported subroutine F 04 AMF) of secondary structure percentages, based on the method of Hennessey and Johnson [14], using their five basis spectra. Different classes of β-turns and β-sheets given by the program were combined to one class of turns and one class of sheets, respectively.

Isolation and characterization of messenger RNA

Total cytoplasmic RNA was isolated from 100 duck lenses as described by Palmiter [15], with slight modifications, and was applied to a column of oligo(dT)-cellulose (Collaborative Research, T2 grade) for the selection of poly(A)-containing RNA. 10 µg of total poly(A)-containing RNA was electrophoretically fractionated on a 1.5% agarose gel, containing 10 mM methylmercury hydroxide, according to the procedure previously described by Bailey and Davidson [16]. In a parallel lane RNA size markers of 2050 (18S), 1600 (16S) and 1300 (14S) nucleotides were run. After electrophoresis, the gel region containing the 1100–2050-nucleotide mRNA was cut in 3-mm slices; the protein-synthesizing activity of recovered mRNA was monitored by *in vitro* translation in a nuclease-treated rabbit reticulocyte lysate [17], followed by SDS-PAGE and autoradiography.

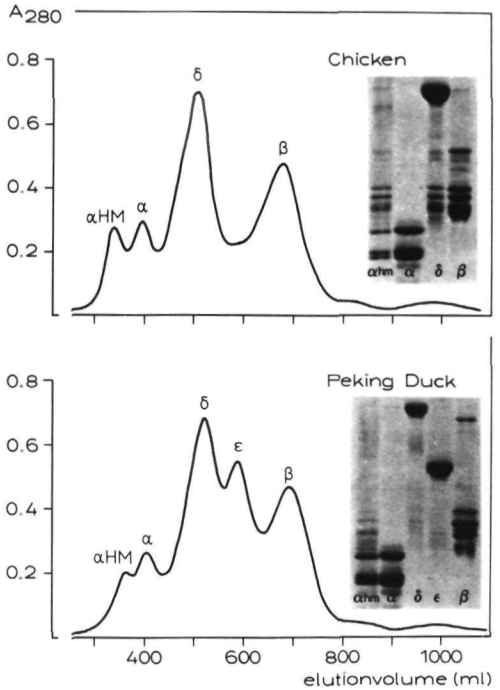


Fig. 1. *Gel filtration of chicken and Peking duck water-soluble lens proteins on Ultrogel AcA 34.* Peak fractions were analyzed by SDS-gel electrophoresis (inserts). The Peking duck β-crystallin fraction also contains the previously reported 48-kDa crystallin [8]

RESULTS

Isolation, gel electrophoresis and immunological characterization of Peking duck ε-crystallin

Comparison of the gel filtration patterns of chicken and Peking duck lens extracts shows an additional peak in the duck chromatogram (Fig. 1). The protein contained in this fraction, designated as ε-crystallin, has a native M_r of 120000, as was determined by Bindels et al. [13], and an apparent subunit M_r of 38000 (Fig. 2A). Two-dimensional gel electrophoresis shows that there is only a single type of ε-crystallin subunit, focusing in the pH-7.5 region (Fig. 2B). It appears, however, that for unknown reasons, charge heterogeneity of the ε-subunit readily appears in many samples, as can be seen in Fig. 3B. This may be due to rapid, multiple deamidations, as is also known to occur, albeit at lower rates, in other crystallines [1]. A similar charge heterogeneity, ranging between pI 6 and 7, was observed on isofocusing of isolated native ε-crystallin, a major component being present approximately at pH 6.7 (results not shown).

Antibodies were raised against Peking duck ε-crystallin, which was purified by preparative SDS-PAGE and electroelution. This antiserum shows no cross-reaction with any other duck lens polypeptide in the immunoblotting procedure (Fig. 3C), nor did any reaction occur with chicken soluble lens proteins (not shown). Moreover, it was not possible to demonstrate an immunological reaction with blotted soluble proteins from other duck organs, such as heart, liver, pectoral muscle, pancreas, brain, ileum and cornea.

After repeated gel filtration the Peking duck ε-crystallin was over 97% pure, as determined by densitometry (Fig. 2A, lane 3). This material was used for further studies of primary and secondary structure, because the remaining slight impurities should not substantially influence the various results.

Primary structure analyses

The amino acid compositions of Peking duck and alligator ε-crystallin, purified by repeated gel filtration, are given in Table 1. Compared to the other crystallins they show a remarkably high value for valine and a very low phenylalanine content.

To enable further comparisons with other crystallins and to facilitate future cDNA sequence analysis, we performed partial primary structure studies of Peking duck ε-crystallin. Dansylation failed to reveal the N-terminal residue, suggesting that ε-crystallin is N-terminally blocked, like α- and β-crystallins. The presence in high yield of a dipeptide Asp-Met in tryptic digests of total ε-crystallin suggests this to be the C-terminal sequence.

We attempted to purify as many tryptic peptides of ε-crystallin as possible, regardless of their order in the chain. To this end the mixture of CNBr fragments of ε-crystallin was fractionated over a Sephadex G-50 sf column. The eluate was pooled in five fractions and each fraction was digested with

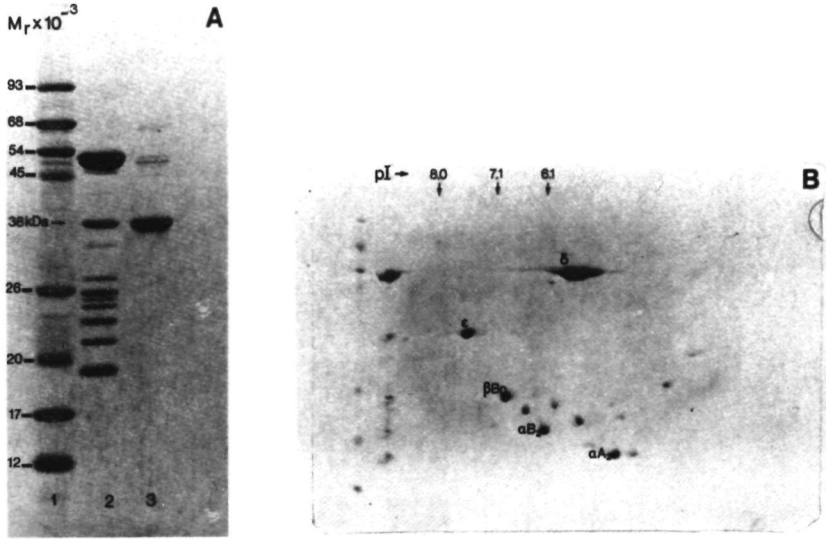


Fig. 2. *One-dimensional (A) and two-dimensional (B) gel electrophoresis of lens proteins.* (A) SDS-gel electrophoresis of (1) marker proteins: cytochrome *c* (12400), myoglobin (17000), α-crystallin A (20000), chymotrypsinogen A (26000), ovalbumin (45000), leucine amino peptidase (54000), bovine serum albumin (68000) and phosphorylase A (93000); (2) Peking duck water-soluble lens proteins; (3) Peking duck ε-crystallin. (B) Two-dimensional gel electrophoresis of Peking duck lens extract. The focusing gel contained 1.3% ampholine (pH 3.5–10) and 3.6% ampholine (pH 5–7). The ε-crystallin subunit focuses in the pH-7.5 region, as was determined by comparison with calf lens polypeptides of known isoelectric points [1]

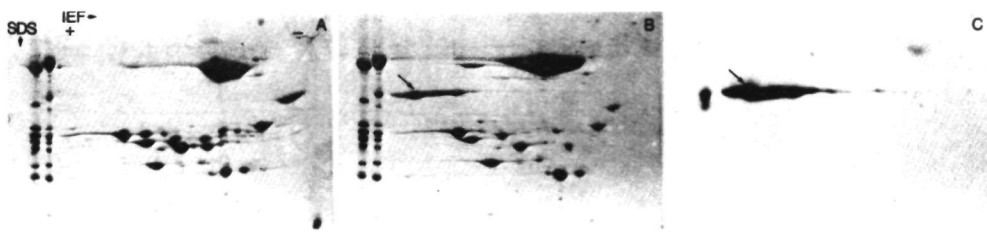


Fig. 3. *Two-dimensional gel electrophoresis of water-soluble lens proteins of (A) chicken and (B) Peking duck; (C) immunoblot of the duck lens extract electropherogram with antiserum against ε-crystallin.* Arrows indicate the ε-crystallin subunit. Focusing gels contained 1% ampholine (pH 3.5–10); 1.3% ampholine (pH 7–9) and 2.6% ampholine (pH 6–8). Reference lanes contained lens extracts of chicken and Peking duck, respectively

trypsin. The digests were either fractionated over Sephadex G-50 sf before peptide mapping or directly subjected to peptide mapping. In this way a total of 38 peptides could be purified (Table 2), accounting for 210 of the approximately 330 residues expected in the ε-crystallin chain.

The sequences of eight tryptic peptides were determined by dansyl-Edman degradation and subdigestions (Fig. 4). These 67 residues account for approximately 20% of the ε-crystallin sequence. They were compared by computer analysis, based on the matrix method of Gibbs and McIntyre [18], with the other known sequences of α-, β-, γ- and δ-crystallins [1–3, 19]. No significant and consistent similarities could be detected.

Secondary and quaternary structure

Information about the secondary structure of ε-crystallin was obtained by comparative CD spectroscopy of duck α-, β-, δ- and ε-crystallin. The CD spectrum of ε-crystallin in the short-wavelength range is presented in Fig. 5, while the secondary structure values of all major duck lens proteins are given in Table 3. These latter results are in accordance with previous reports [2, 20]. It is obvious that ε-crystallin, like δ-crystallin, but in contrast to α- and β-crystallin, has a relatively high amount of α-helical structure.

The fact that the M_r of native Peking duck ε-crystallin, as determined by high-performance gel permeation chroma-

Table 1. *Amino acid compositions of ε-crystallin from Peking duck and alligator, compared with literature values for other crystallins*
Values for ε-crystallins are the average of duplicate analyses after 24 h, 48 h and 72 h of hydrolysis. Values for threonine and serine are obtained by extrapolation to zero-time hydrolysis; valine and isoleucine have values for 72-h hydrolysis. Crystallins used for comparison were bovine α-, β_{Low}- and γ-crystallins [1], chicken δ-crystallin [2] and lamprey 48-kDa crystallin [8]. All values are residues/1000 residues

Amino acid	ε		Comparisons				
	duck	alligator	α	β_{Low}	γ	δ	48-kDa
Asp	98	102	86	85	116	71	111
Thr	58	48	34	30	22	76	39
Ser	83	86	103	81	42	96	54
Glu	98	102	105	146	131	130	105
Pro	33	37	81	59	58	23	38
Gly	81	77	60	91	93	56	91
Ala	75	62	44	49	32	80	99
Cys	23	19	5	9	42	3	16
Val	114	108	58	63	48	80	69
Met	18	23	12	10	37	8	20
Ile	47	56	47	35	42	74	65
Leu	116	113	87	62	81	151	76
Tyr	21	23	31	41	76	8	31
Phe	14	17	76	43	48	22	37
His	31	32	39	45	37	12	24
Lys	80	82	48	54	24	72	93
Arg	35	40	73	61	111	39	45

tography in combination with low-angle laser light scattering, is found to be 120000 [13], together with the subunit M_r of 38000, implies a trimeric quaternary structure of identical subunits.

Isolation and characterization of messenger RNA

The poly(A)-containing RNA from Peking duck lenses was isolated and electrophoretically fractionated as described in Materials and Methods. The gel region containing the mRNAs between approximately 1100 and 2050 nucleotides in length was cut in 3-mm slices. The mRNA was recovered from each of the slices and translated in a rabbit reticulocyte system in the presence of [^{35}S]methionine. The translation products were analyzed by SDS-PAGE and autoradiography (Fig. 6). The size of the ε-crystallin mRNA was estimated to be approximately 1450 nucleotides, as deduced from the M_r of the mRNA in the gel fraction which synthesized the highest quantity of ε-crystallin (fraction 14 in Fig. 6). The value of 1450 nucleotides is in good accordance with the finding of a 15-S ε-mRNA from sucrose density gradient centrifugation experiments (data not shown). Since the 38000-M_r ε-crystallin chain must contain approximately 330 amino acids, the coding and non-coding regions of the mRNA comprise about 990 and 450 nucleotides, respectively.

Occurrence in other species

An SDS-electrophoretic comparison of lens extracts from a variety of vertebrate species (Fig. 7) reveals that the ε-sub-

Table 2. *Amino acid compositions of tryptic peptides from CNBr fragments of Peking duck ε-crystallin*
Purification procedures were (A) gel filtration over Sephadex G-50 sf, (B) peptide mapping (high-voltage paper electrophoresis at pH 6.5 followed by descending chromatography) and (C) reelectrophoresis at pH 3.8 of neutral peptides. Charges of peptides were estimated from their electrophoretic mobilities at pH 6.5. No correction was made for hydrolytic destruction of threonine and serine. Homoserine plus homoserine lactone (Hse) could not be determined quantitatively. Low values for valine and isoleucine in peptides 5 and 7 are probably due to incomplete cleavage of peptide bonds after 22 h of hydrolysis. The presence of tryptophan was deduced from ultraviolet fluorescence after peptide mapping. Amino acid analyses shown for peptides 9, 19 and 20 were obtained after a 72-h hydrolysis. Peptide 10 is peptide 3 plus Lys. Peptide 17 is peptide 16 plus Leu-Hse. Peptide 33 is found as Asp-Met on peptide maps after tryptic digestion of native ε-crystallin. The number of residues given at the bottom is the most probable number in the analyzed peptides, i.e. giving the nearest to integral values (taking into account hydrolytic destruction and incomplete hydrolysis). Excluding peptides 3, 6, 6–7, 7, 8 and 16 (which result from further cleavage of peptides 10, 6–7–8 and 17), this table accounts for 210 of the approximately 330 residues in the ε-crystallin chain

Peptide	1	2	3	4	5	6	6–7	6–7–8	7	8	9	10	11
Asp		2.8	2.2	1.8	1.8	1.9	3.0	2.8	1.0			2.0	1.1
Thr		1.0			0.9	0.9	1.0	0.9					1.0
Ser	0.9	1.0			0.9			1.0		0.7	1.0		1.8
Glu	3.8	1.0	1.8								2.9	1.7	1.1
Pro													
Gly	1.2	1.1			2.0								
Ala			1.1		1.0	2.2	2.9	2.8	1.0		1.2	1.1	1.0
Cys													
Val			1.0		0.6	1.0	1.5	1.4	0.5		2.8	1.0	
Hse													
Ile					0.5		0.5	0.4	0.4		0.9		1.0
Leu			2.0	0.9	0.9							1.8	1.1
Tyr						1.0	0.9	1.0			1.0		
Phe													
His													
Lys		2.1	1.9	1.2			1.2	2.0	1.1	1.3		3.3	1.0
Arg	1.1				1.0						1.0		
Trp		+											+
Purif.	AB	B	AB	AB	AB	AB	AB	ABC	AB	B	AB	AB	ABC
Residues	7	10	10	4	11	7	12	14	5	2	11	11	10
Charge	−1	−1	−1	−1	−1	−1	−1	0	0	+1	−1	0	0

Table 2 (Continued)

Peptide	12	13	14	15	16	17	18	19	20	21	22	23	24
Asp	1 0									0 7	1 0	1 0	
Thr					0 9	1 0		0 8				0 9	1 1
Ser					2 7	2 8						1 0	2 6
Glu		1 2	1 0				1 0		1 0		1 0	2 2	
Pro					1 9	1 9			0 9				
Gly		1 0						1 0	1 0	1 2		0 9	
Ala			0 8		4 2	4 0		1 0					
Cys													
Val					2 1	2 3		4 0	1 0	1 9	1 0		3 1
Hse		+		+		+							
Ile								0 9	2 5				
Leu	1 0	0 8		1 1		1 0					2 0	2 8	2 3
Tyr				1 0									
Phe										0 8		0 8	
His												2 0	1 0
Lys	1 0	0 9			1 2	1 2	1 0		1 0	1 3		1 2	1 9
Arg			1 2					1 1			1 0		
Trp													
Purif	ABC	ABC	BC	AB	B	B	BC	B	AB	B	B	B	B
Residues	3	5	3	3	13	15	2	9	8	6	6	13	12
Charge	0	0	0	0	+1	+1	0	+1	+1	+1	+1	+1	+3

Peptide	25	26	27	28	29	30	31	32	33	34	35	36
Asp									1 2	1 0		
Thr						1 0						
Ser						2 0	0 9	2 0			1 0	
Glu				0 9	0 9		0 8					1 2
Pro								1 0			1 0	
Gly											1 0	
Ala								1 0				0 9
Cys												
Val					1 0	2 9		1 0				
Hse									0 8			
Ile							1 1					
Leu		1 0				1 1				1 0	1 8	1 0
Tyr												0 7
Phe	0 9											
His			1 0		1 0	0 9						
Lys		1 0		1 1	1 1	1 1	1 2	1 0			1 0	
Arg	1 1		1 0							1 0		1 2
Trp												
Purif	B	B	B	B	AB	AB	B	B	AB	AB	AB	ABC
Residues	2	2	2	2	4	9	4	6	2	3	6	5
Charge	+1	+1	+2	+1	+1	+2	+1	+1	−1	0	+1	0

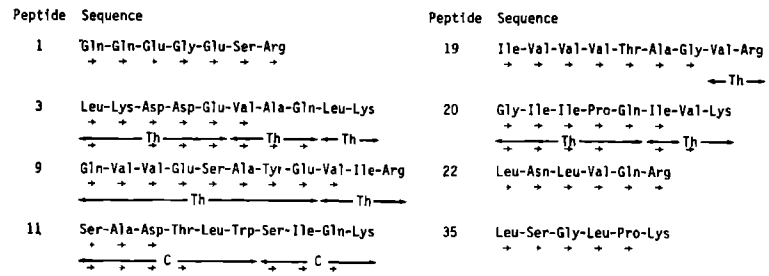


Fig 4 *Amino acid sequences of eight γ-crystallin tryptic peptides* Peptides were sequenced by dansyl-Edman degradation (→) Subdigestions were carried out with thermolysin (Th) and chymotrypsin (C)

unit is only found in certain avian and reptilian lenses. Interestingly, among the investigated reptiles ε-crystallin is only present in the crocodilians, which are supposed to be the closest living relatives of the birds [21]. ε-Crystallin is especially abundant in ducks and geese, but is also present in considerable amounts in the lenses of heron, some waders and gulls. It also appeared to be a major compound in the lenses of gannet and buzzard (not shown). A 37500-M_r component in *Rana esculenta* lens extract, probably comparable to the previously reported *Rana pipiens* polypeptide [22], showed no reaction in the immunoblotting procedure with the ε-antiserum.

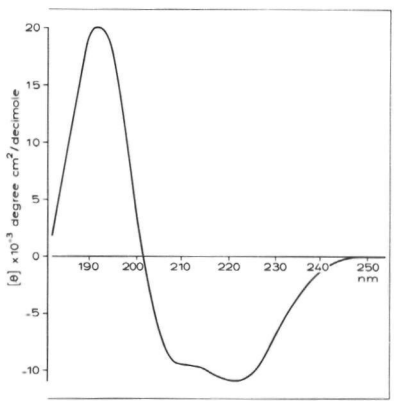


Fig. 5. *Far-ultraviolet CD spectrum of Peking duck ε-crystallin.* The spectrum is the average of 11 scans. Path length is 0.1 mm, protein concentration is 1 mg/ml

Interspecies differences in ε-crystallin

To obtain some information about the evolution rate of ε-crystallin, a number of tryptic peptides was isolated, in the same way as described above, from ε-crystallin of the tufted duck (belonging to a different family in the same order, Anseriformes, as the Peking duck), the heron (representing a different avian order, the Ciconiiformes) and the alligator (as representative of the class Reptilia). Comparison of amino acid compositions of homologous peptides from these ε-crystallins with those of Peking duck (Table 4) provides a preliminary indication of the extent of evolutionary divergence of the ε-crystallin sequences among birds and reptiles. 23 homologous peptides (totalling 143 residues) of tufted duck and Peking duck showed not a single substitution. 20 homologous peptides (127 residues) of the heron contained 8 substitutions, and 17 peptides (96 residues) isolated from alligator showed 6 substitutions as compared to the homologous Peking duck peptides.

Because the compared ε-crystallin peptides of alligator, heron and ducks represent only 30–43% of the total chain and since amino acid substitutions are known to be often unequally distributed over the length of a chain, the observed differences may not be representative for ε-crystallin as a

Table 3. *Secondary structure values of the major Peking duck water-soluble eye lens proteins*

Structure	Amount in crystallin			
	α	β	δ	ε
	%			
α-Helix	14 ± 1	11 ± 1	50 ± 5	30 ± 1
β-Sheet	27 ± 1	47 ± 1	26 ± 3	23 ± 1
β-Turn	19 ± 1	19 ± 1	15 ± 1	15 ± 1
Random coil	40 ± 1	23 ± 1	9 ± 3	32 ± 1

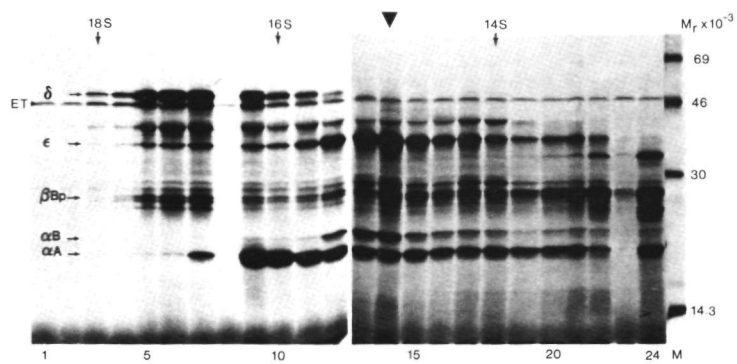


Fig. 6. *Analysis by SDS-PAGE of the ^{35}S-labeled translation products of duck lens mRNA fractions.* The successive lanes of this autoradiograph contain the translation products of the mRNAs extracted from the respective slices of a denaturing electrophoresis gel on which total poly(A)-containing RNA had been fractionated. The fractions corresponding with the 18-S, 16-S and 14-S RNA size markers on the denaturing gel are indicated. The arrow above lane 14 indicates the translation peak of the ε-crystallin mRNA, corresponding to a size of approximately 1450 nucleotides. Lane M contains the [^{14}C]methylated marker proteins: lysozyme (14.3 kDa), carbonic anhydrase (30 kDa), ovalbumin (46 kDa) and serum albumin (69 kDa). ET: endogenous translation product of the reticulocyte lysate

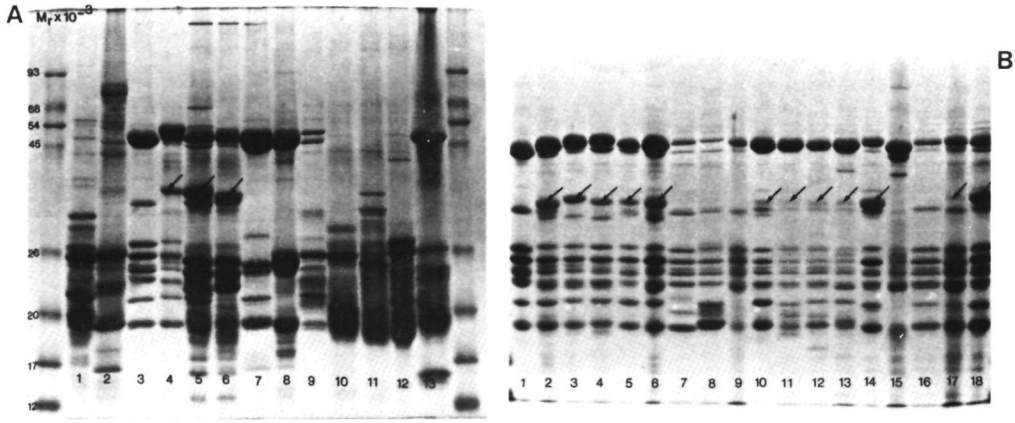


Fig. 7. *SDS-gel electrophoretic analysis of (A) 13 vertebrate and (B) 18 avian lens extracts.* (A) (1) Calf (*Bos taurus*), (2) echidna (*Tachyglossus aculeatus*), (3) chicken (*Gallus domesticus*), (4) Peking duck (*Anas platyrhynchos*), (5) alligator (*Alligator mississippiensis*), (6) caiman (*Caiman crocodilus*), (7) cape monitor (*Varanus exanthematicus*), (8) western diamond-back rattlesnake (*Crotalus atrox*), (9) red-eared turtle (*Chrysemys scripta elegans*), (10) clawed toad (*Xenopus laevis*), (11) cichlid (*Tilapia mossambica*), (12) spiny dogfish (*Squalus acanthias*), (13) lamprey (*Petromyzon marinus*). Arrows indicate the ε-crystallin subunit, as identified by immunoblotting. (B) (1) Chicken, (2) bean goose (*Anser fabilis*), (3) gadwall (*Anas strepera*), (4) common golden-eye (*Bucephala clangula*), (5) mute swan (*Cygnus olor*), (6) tufted duck (*Aythya fuligula*), (7) woodpecker (*Dendrocopos major*), (8) cuckoo (*Cuculus canorus*), (9) coot (*Fulica atra*), (10) yellow-billed loon (*Gavia adamsii*), (11) red-throated loon (*Gavia stellata*), (12) common loon (*Gavia immer*), (13) arctic loon (*Gavia arctica*), (14) heron (*Ardea cinerea*), (15) gentoo penguin (*Pygoscelis papua*), (16) fulmar (*Fulmarus glacialis*), (17) curlew (*Numenius arquata*), (18) oystercatcher (*Haematopus ostrelagus*). Arrows indicate the ε-crystallin subunit, as identified by immunoblotting

Table 4. *Homologous tryptic peptides in ε-crystallin of Peking duck (Du), heron (He) and alligator (Al)*

Peptide numbers correspond to those in Table 2 and sequences (if known) are as given in Fig. 4. The one-letter notation for amino acids is used, and the recommended punctuation and brackets for partly known sequences is followed [26]. None of the heron or alligator peptides has been sequenced. The order of fragments in peptide 6–7–8 is based on the finding of peptides 6–7 and 7–8. Peptide 33 is the putative C-terminus of the total ε-chain

	1	2	3
Du	Q-Q-E-G-E-S-R	B,B,B,T,S,Z,G,W,K,K	L-K-D-D-E-V-A-Q-L-K
He		—T—	
Al			—R

	6-7-8	9
Du	(D,N,T,A,A,V)Y(D,A,V,I)K-S-K	Q-V-V-E-S-A-Y-E-V-I-R
He		
Al		

	11	12	13	14	15
Du	S-A-D-T-L-W-S-I-Q-K	D,L,K	E,G,K,L,M	E,A,R	L,Y,M
He		—V—		—S—	
Al					

	17	18	19
Du	L-M(T,S,S,S,P,P,A,A,A,A,V,V)K	E-K	I-V-V-V-T-A-G-V-R
He	—N,N,G—		
Al			—H—

	21	22	25	26	27	28
Du	N,G,V,V,F,K	L-N-L-V-Q-R	F-R	L-K	H-R	S-K
He	—N—					
Al	—N—					

	29	30	33	35
Du	Q,V,H,K	T,S,S,V,V,V,L,H,K	D-M	L-S-G-L-P-K
He			—L—	
Al		—P— —M—	—L—	

whole. It is nevertheless of interest to see whether the findings for these parts of the ε-crystallin chain would indicate the same or a different rate of evolution as compared to the other crystallins.

Considering that the avian orders diverged approximately 70×10^6 year ago, and the birds and crocodiles for 200×10^6 years, one can estimate, as outlined earlier [9], that the investigated peptides of ε-crystallin evolved at a rate between 2 and 4 substitutions per 100 residues per 100 million years. This would be comparable to the rate of evolution of α-crystallin (3% sequence change per 100×10^6 years [7]), and β-crystallins (4% change per 100×10^6 years [19, 23]). γ-Crystallins (7% change per 100×10^6 years, as can be calculated from the results presented in [3] or 10% as estimated in [23]) and δ-crystallins seem to evolve somewhat faster (10% change per 100×10^6 years [9]).

DISCUSSION

The results of gel filtration and electrophoretic analyses reveal the presence of a 120000-M_r water-soluble protein, composed of three identical subunits with an apparent M_r of 38000 and a pI value of about 7.5 in Peking duck lenses. This protein makes up to about 10% of the water-soluble lens proteins in the duck. Immunoblotting results indicated that this protein, ε-crystallin, is not immunologically related to any other crystallin, nor to soluble proteins extractable from a variety of duck organs. Also the amino acid composition, the sequence results, its quaternary structure and secondary structure characteristics support the conclusion that ε-crystallin represents a novel class of crystallins.

It should nevertheless be noted that ε crystallin has some features in common with δ-crystallin Both proteins only occur in birds and reptiles they have a relatively high α helical content and possibly related herewith a low value for proline and a rather high leucine and lysine content In contrast to the monomeric 48 kDa and , crystallins and the heterogenous aggregates of α and β crystallins they show a well defined trimeric or tetrameric quaternary structure respectively An interesting difference is that δ crystallin is a typically embryonic lens protein [2] while ε crystallin is the latest lens protein to appear during development of the duck lens [24]

The possibility that ε crystallin is somehow a post translational derivative of δ-crystallin is ruled out by the absence of immunological or sequence relationship and most conclusively by the isolation of a separate mRNA for ε crystallin

It is tempting to suggest that the similarities between ε and δ-crystallin reflect specific requirements for the proper functioning of lens proteins in the avian and reptilian lenses which are characterized by a very soft consistency and great plasticity in relation to their unique accomodative properties [7]

It is remarkable that the major lens proteins in the duck and in many other birds and reptiles are the products of five unrelated gene families α β ,- δ- ε- and 48 kDa crystallins [1 8] while in mammalian lenses only two gene families (α and β ,) take care of the crystallin production [25] The scattered occurrence of ε-crystallin among avian species is also intriguing While it is found in considerable amounts in some species like duck it is apparently absent in others like chicken This opens interesting possibilities for comparative studies of gene expression

We thank Drs H C Dessauer J Wittel Th Smit Ms T Prins and the Central Animal Facilities of the University of Nijmegen School of Medicine for making available eye lenses for this investigation Marlies Versteeg for performing amino acid analyses Jack Leunissen for comparative computer analyses and Mirjam Dubelaar for assistance in developing of CD computer programs This investigation was supported by the Foundation for Fundamental Biological Research (BION) which is subsidized by the Netherlands Organization for Pure Research (ZWO)

REFERENCES

1 Bloemendal H (ed) (1981) *Molecular and cellular biology of the eye lens* John Wiley & Sons New York
2 Piatigorsky J (1984) *Mol Cell Biochem 59* 33–56
3 Moormann R J M (1984) *Structure and genetic organization of the rat crystallin gene family* Thesis University of Nijmegen
4 Kondoh H Yasuda K & Okada T S (1983) *Nature (Lond) 270* 406–411
5 Clayton R M (1974) in *The eye* (Davson H & Graham L T eds) pp 399–494 Academic Press London
6 Driessen H P C Herbrink P Bloemendal H & de Jong W W (1981) *Eur J Biochem 121* 83–91
7 De Jong W W (1981) in *Molecular and cellular biology of the eye lens* (Bloemendal H ed) pp 221–278 John Wiley & Sons New York
8 Stapel S O & de Jong W W (1983) *FEBS Lett 162* 305–309
9 De Jong W W Stapel S O & Zweers A (1981) *Comp Biochem Physiol 69B* 593–598
10 Laemmli U K (1970) *Nature (Lond) 227* 680–685
11 O Farrell P H (1975) *J Biol Chem 250* 4007–4021
12 Burnette W N (1981) *Anal Biochem 112* 195–203
13 Bindels J G Bessems G J J de Man B & Hoenders H J (1983) *Comp Biochem Physiol 76B* 47 55
14 Hennessey J P Jr & Johnson W C Jr (1981) *Biochemistry 20* 1085–1094
15 Palmiter R D (1974) *Biochemistry 13* 3606–3615
16 Bailey J M & Davidson N (1976) *Anal Biochem 70* 75–85
17 Pelham H R B & Jackson R J (1976) *Eur J Biochem 67* 247–256
18 Gibbs A J & McIntyre G A (1970) *Eur J Biochem 16* 1–11
19 Berbers G A M Hoekman W A Bloemendal H de Jong W W Kleinschmidt T & Braunitzer G (1984) *Eur J Biochem 139* 467–479
20 Horwitz J (1976) *Exp Eye Res 23* 471–481
21 Romer A S (1966) *Vertebrate Paleontology* 3rd ed University of Chicago Press Chicago
22 Zigler J S & Sidbury J B (1976) *Comp Biochem Physiol 55B* 19–24
23 Tomarev S I Zinovieva R D Chalovka P Krayev A S Skryabin K G & Gause G G Jr (1984) *Gene 27* 301–308
24 Brahma S K & van der Starre H (1983) *Curr Eye Res 2* 663–677
25 Harding J J & Dilley K J (1976) *Exp Eye Res 22* 1–73
26 IUPAC IUB Commission on Biochemical Nomenclature (1984) *Eur J Biochem 138* 9–37
27 Tomarev S I Zinovieva R D Dolgilevich S M Luchin S V Krayev A S Skryabin K G & Gause G G Jr (1984) *FEBS Lett 171* 297–302
28 Stapel S O Zweers A Versteeg M Dodemont H & de Jong W W (1982) *Proceedings 5th International Congress of Eye Research* p 43 Koningshof Veldhoven

S O Stapel A Zweers and W W de Jong Laboratorium voor Biochemie Universiteit van Nijmegen Geert Grooteplein Noord 21 NL 6525 EZ Nijmegen The Netherlands

H J Dodemont Department of Neurobiology Weizmann Institute of Science P O Box 26 II 76100 Rehovot Israel

J H Kan Laboratorium voor Biochemie Rijks Universiteit van Leiden Wassenaarseweg 64 NL 2333 AL Leiden The Netherlands

Note added in proof (received December 20 1984) After submission of this paper the name ε crystallin has also been proposed for a novel 35 kDa lens protein from the frog *Rana temporaria* of which protein the partial cDNA sequence has been determined [27] Apart from the similarity in subunit molecular mass and their oligomeric structure the presently available data do not seem to support an evolutionary relationship between this frog crystallin and the avian and reptilian ε-crystallin described in the present paper and mentioned earlier [13 24 28] Since the sequence data for both proteins are still incomplete a final decision about this relationship and the consequent nomenclature should await more complete sequence information

Lamprey 48-kDa lens protein represents a novel class of crystallins

Steven O. Stapel and Wilfried W. de Jong

Laboratorium voor Biochemie, Universiteit van Nijmegen, Geert Grooteplein Noord 21, 6525 EZ Nijmegen, The Netherlands

Received 30 August 1983

SDS-PAGE revealed a major M_r 48000 polypeptide of pI around 8 in the water-soluble fraction of lamprey lenses. It occurs as a monomeric protein, and its amino acid composition and tryptic peptides show no resemblances to α-, β-, γ- or δ-crystallin. Immunoblotting with antiserum against the 48-kDa protein revealed an immunologically related polypeptide of similar M_r in reptiles, several birds and a fish, but showed no cross-reactivity with any other water-soluble lens component. The 48-kDa protein is not detected in many birds and fishes, and in the investigated mammals and amphibians.

Crystallin *Lens protein* *Lamprey* *Protein evolution* *Immunoblotting*

1. INTRODUCTION

The vertebrate eye lens contains a number of organ-specific water-soluble proteins, the crystallins, which are classically divided into different groups, designated as α-, β-, γ- and δ-crystallins [1]. Especially the δ-crystallins have caught the investigators eye, not only by the circumstance that they show a number of features concerning regulation of protein synthesis, protein structure and evolutionary development which sets them apart from the other crystallins [2–4], but also by the fact that they appear only in the lenses of the sauropsidan species (birds and reptiles), although cloned chicken δ-crystallin cDNA sequences seem to hybridize weakly to genomic DNA from other phylogenetic groups, insects included [5]. δ-Crystallin is a tetrameric protein, ranging in M_r from 150000 to 200000 and is composed of subunits with M_r between 45000 and 50000 [6–8]. This clearly distinguishes δ-crystallin from the other crystallins, which have monomeric M_r-values between 20000 and 34000 [1]. We therefore were surprised to find, in the course of a comparative electrophoretic analysis of vertebrate lens extracts in SDS-gels, a major M_r 48000 polypeptide in the eye lenses of sea lamprey, and it seemed worthwhile to isolate and characterize this component in order to establish its possible relationship to sauropsidan δ-crystallin.

2. MATERIALS AND METHODS

Sea lampreys (*Petromyzon marinus*) were caught in the river Maas, and after dissection the eye lenses were stored at −20°C. Lenses of 12-week old chickens and of other species were obtained and treated as in [4]. Due to their extreme hardness, the lamprey lenses had to be homogenized by means of a Polytron apparatus (Kinematica GMBH, Luzern) in a small volume of 1% ammonium bicarbonate, pH 7.9. After centrifugation for 30 min at 10000 × g, the supernatant was applied to a column (125 × 3.5 cm) of Ultrogel AcA-34 (LKB) and eluted at room temperature with the solvent mentioned. After gel filtration, fractions containing protein material were pooled an lyophilized. Slab gel electrophoresis was performed in 13% polyacrylamide gels containing 0.1% sodium dodecylsulfate as in [9]. Two-dimensional gel electrophoresis was performed as in [10]. Peptide mapping and amino acid analysis of pro-

Published by Elsevier Science Publishers B V
00145793/83/$3.00

teins and their tryptic peptides were performed as in [4]. To obtain antisera against the sea lamprey protein and chicken δ-crystallin, immunization was carried out with protein samples which, after gel filtration, had been purified by preparative SDS-PAGE and were removed from the gel material by electroelution. Rabbits were injected with 2 mg of protein in the presence of complete Freund's adjuvant, and blood samples were taken after 3 weeks. Electroblot transfer and immunoautoradiography with [^{125}I]protein A were performed as in [11].

3. RESULTS AND DISCUSSION

The sea lamprey water-soluble lens extract contains a major polypeptide component of M_r 48000, which is in the molecular mass range of the sauropsidan δ-crystallin subunits (fig.1A). However, its elution volume on a gel filtration column, calibrated with calf lens crystallins of known M_r, shows that it behaves like a monomer, in contrast to the tetrameric δ-crystallin (fig.2). Two-dimensional gel electrophoresis of chicken and lamprey lens extracts shows that the 48-kDa protein focuses

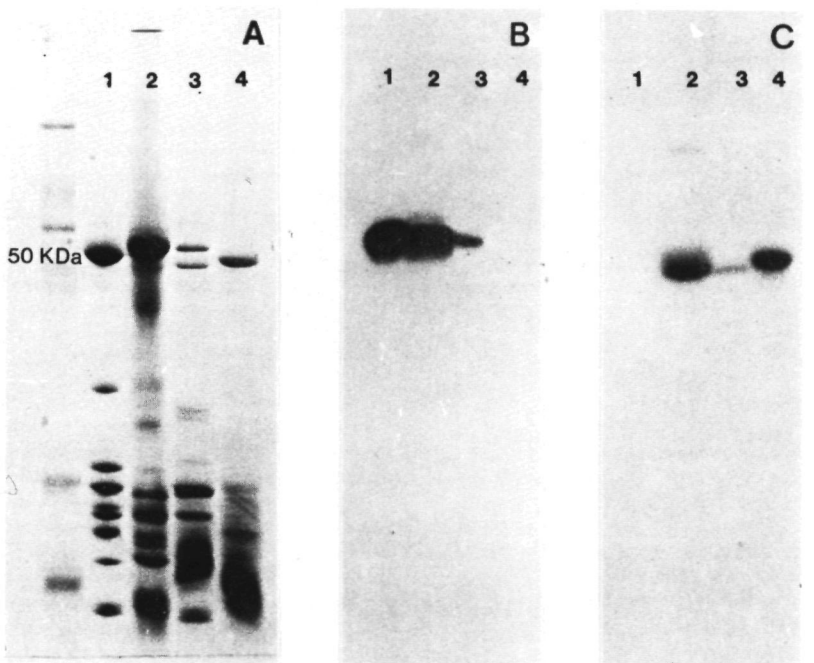


Fig.1(A). SDS-PAGE of water-soluble lens proteins of (1) chicken (*Gallus domesticus*); (2) alligator (*Alligator mississippiensis*); (3) turtle (*Pseudemys scripta-elegans*); (4) sea lamprey (*Petromyzon marinus*); (B) Immunoblotting of the same gel with antiserum against chicken δ-crystallin; (C) Immunoblotting with antiserum against lamprey 48-kDa protein.

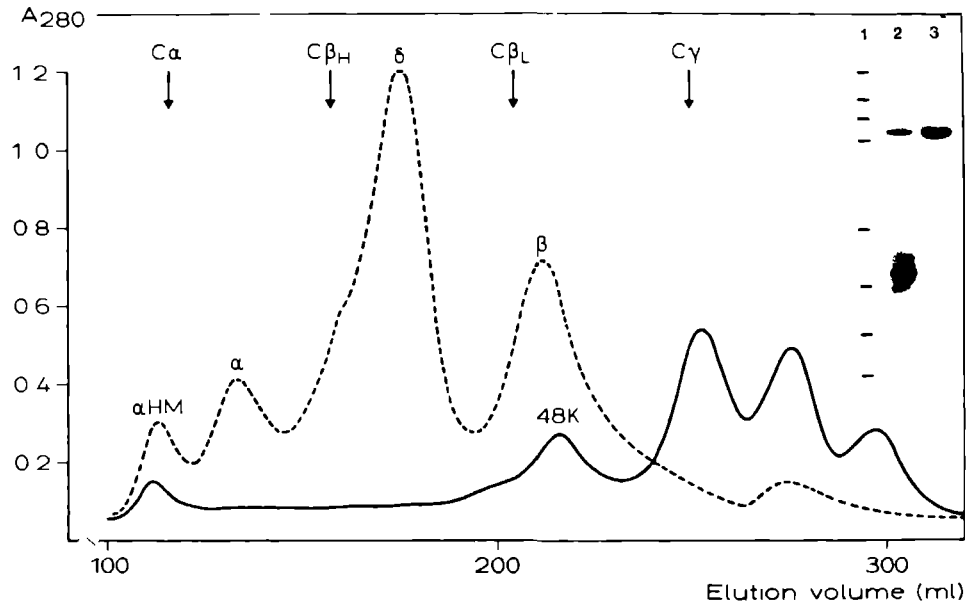


Fig 2 Comparative gel filtration-on Ultrogel AcA-34 of lens extract of chicken (- - -) and sea lamprey (——) Arrows indicate the elution volumes of calf α-crystallin (M_r 800000), β_{High}-crystallin (M_r 180000), β_{Low}-crystallin (M_r 50000) and γ-crystallin (M_r 20000) Insert SDS-PAGE of (1) marker proteins cytochrome *c* (12400), myoglobin (17000), α-crystallin A (20000), chymotrypsinogen A (26000), ovalbumin (45000), leucine aminopeptidase (54000), bovine serum albumin (68000) and phosporylase A (93000), (2) lamprey lens extract, (3) 48-kDa protein

at an isoelectric point of about 8, while the δ-crystallin subunits are found in the pH 5–6 region (fig 3A,B) Amino acid analysis of the 48-kDa protein (Table 1) does not, amongst others, reveal the low value for tyrosine, and the high values for threonine and leucine which are characteristic for δ crystallins [4,7] Peptide mapping followed by amino acid analyses of 22 tryptic peptides, accounting for about 30% of the total polypeptide chain, did not reveal any significant resemblances to δ-crystallin peptides either

Immunoblotting of the lens extracs shown in fig 1A, using antiserum against chicken δ-crystallin, gives the expected reaction with the δ-crystallin bands of chicken, alligator and turtle, but not with any of the lamprey polypeptides (fig 1B) On the other hand, the antiserum against the 48 kDa protein not only reacts with the lamprey component, but also evokes a similar reaction with a band just below δ-crystallin in the turtle and alligator lens extracts (fig.1C).

In the course of further comparative immunoblotting a number of vertebrate water-soluble lens extracts were studied, including mammals (calf, horse, rhesus monkey, and the Australian spiny anteater), amphibians (*Rana esculenta, Xenopus laevis*), reptiles (caiman, rattlesnake, monitor lizard), birds (emu, penguin, gannet, peking duck, buzzard, coot, gull, pigeon, budgerigar, cuckoo and eagle owl), fishes (river lamprey, dogfish, alligator gar, cichlid and carp). A clear immunological reaction appeared with a ±48-kDa band in lens extracts of all reptiles, a number of avian species (emu, penguin, gannet, duck, gull, cuckoo) and with the lens extracts of alligator gar and river lamprey. The anti-48-kDa-serum did not react with any other protein band of any lens extract subjected to the immunoblotting

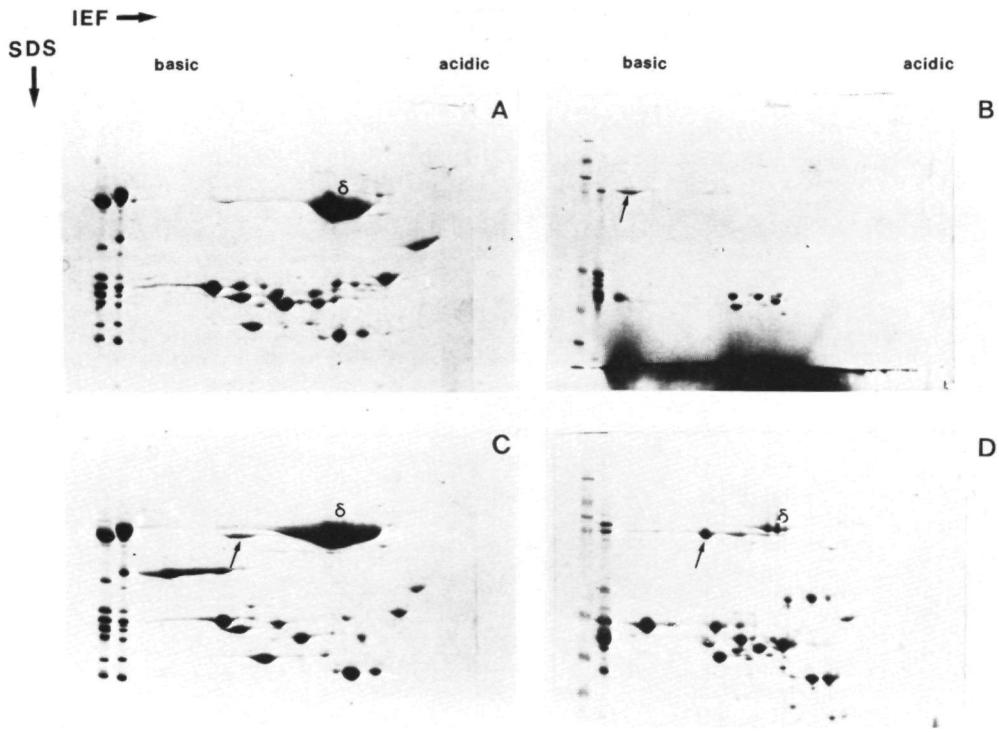


Fig.3. Two-dimensional gel electrophoresis of lens extracts of chicken (A), sea lamprey (B), peking duck (C) and turtle (D), using 1% ampholine (pH 3.5–10); 1.3% ampholine (pH 7–9) and 2.6% ampholine (pH 6–8) in the focusing gels. Arrows indicate the 48-kDa protein in lamprey and the corresponding components in Peking duck and turtle. The turtle component apparently occurs in different charge forms, probably due to deamidation. Reference lanes of fig. 3A and 3C contain lens extracts of chicken and Peking duck, respectively.

procedure (not shown). Comparative gel filtration of lens extracts from alligator, lizard, turtle and peking duck clearly showed the presence of the 48-kDa-related component in the β_{Low}-fractions, which is an indication of its monomeric behaviour in these species. Two-dimensional gel electrophoresis of the lens extracts of Peking duck and turtle (fig.3C,D) pointed out that the protein in birds and reptiles, as compared to the lamprey component, has a lower isoelectric point, of about 6. In fact, the two-dimensional gel of chicken lens extract (fig.3A) also reveals a minor spot in a position corresponding to the duck 48-kDa component. It thus can not be excluded that a small amount of the 48-kDa protein, although not detected by one-dimensional gel immunoblotting, is present in chicken lens as well.

These data allow us to conclude that the lamprey 48-kDa lens protein is not detectably related to δ-crystallin, nor to the α-, β- or γ-crystallins. It represents a new class of crystallins with a scattered distribution among vertebrates. It is immunologically related to the turtle 46-kDa (pI 6.2) polypeptide observed but not further characterized [12].

The presence of the 48-kDa protein in distantly related taxa with structurally very different lenses, such as lampreys, the fish superorder Holostei,

Table 1

Amino acid composition of lamprey 48 kDa lens protein compared with literature values for chicken δ-crystallin [7] and bovine α- β_{Low}- and γ-crystallins [1] (residues/1000 residues)

	48-kDa[a]	δ	α	β_{Low}	γ
Asp	111	71	86	85	116
Thr	39	76	34	30	22
Ser	54	96	103	81	42
Glu	105	130	105	146	131
Pro	38	23	81	59	58
Gly	91	56	60	91	93
Ala	99	80	44	49	32
Cys	16	3	5	9	42
Val	69	80	58	63	48
Met	20	8	12	10	37
Ile	65	74	47	35	42
Leu	76	151	87	62	81
Tyr	31	8	31	41	76
Phe	37	22	76	43	48
His	24	12	39	45	37
Lys	93	72	48	54	24
Arg	45	39	73	61	111

[a] Values are the average of duplicate analyses after 24, 48 and 72 h of hydrolysis Values for threonine and serine are obtained by extrapolation to zero time hydrolysis, valine and isoleucine have values for 72 h hydrolysis

reptiles and several avian species, while the protein is not detected in many other, sometimes closely related taxa, makes it difficult to attribute a specific structural-functional significance to this protein, nor can it be seen as a characteristic phylogenetic trait This protein clearly is the product of an evolutionarily old gene, which has largely or completely been silenced in the eye lenses of the investigated mammals, amphibians, teleost fishes, shark, and in many birds, while it is still expressed in other groups Especially intriguing is the situation among birds, where the 48-kDa component occurs scattered over different orders It would be of interest to establish whether the gene for this protein still occurs in the chromosomal DNA of species which seem to have lost the component in their eye lenses, and if so, to study the structural changes that led to its inactivation

ACKNOWLEDGEMENTS

We are grateful to Marlies Versteeg for performing the amino acid analyses Alligator lenses were kindly provided by Dr H C Dessauer (New Orleans), and most avian lenses by Dr J Wattel (Amsterdam) This investigation was supported by the Foundation for Fundamental Biological Research (BION), and the Netherlands Foundation for Chemical Researsch (SON), which are subsidized by the Netherlands Organization for Pure Research (ZWO)

REFERENCES

[1] De Jong, W W (1981) in Molecular and Cellular Biology of the Eye Lens (Bloemendal, H ed) pp 221–278, Wiley, New York
[2] Shinohara, T and Piatigorsky, J (1977) Nature 270, 406–411
[3] Horwitz, J and Piatigorsky, J (1980) Biochim Biophys Acta 624, 21–29
[4] De Jong, W W , Stapel, S O and Zweers, A (1981) Comp Biochem Physiol 69B, 593–598
[5] Piatigorsky, J (1983) Mol Cell Biochem in press
[6] Rabaey, M , Lagasse, A and De Mets, M (1969) Acta Zool Pathol Antverp 48, 63–71
[7] Piatigorsky, J , Zelenka, P and Simpson, R T (1974) Exp Eye Res 18, 435–446
[8] Williams, L A and Piatigorsky, J (1979) Eur J Biochem 100, 349–357
[9] Laemmli, U K (1970) Nature 227, 680–685
[10] O'Farrell, P H (1975) J Biol Chem 250, 4007–4021
[11] Burnette, W N (1981) Anal Biochem 112, 195–203
[12] Williams, L A , Piatigorsky, J and Horwitz, J (1982) Biochim Biophys Acta 708, 49–56

II-4: ε-crystallin, 48 KDa-crystallin and the grouping of avian orders

The scattered occurrence of "ε-crystallin" and "48 KDa-crystallin" among avian species made it very tempting to test their value as taxonomic trait in a cladistic analysis of avian orders.

Unfortunately, both components turned out to be present before the avian radiation, as could be concluded from the studies of reptilian, mammalian, fish and lamprey lenses.

They must therefore be considered to be "primitive" character states (plesiomorphies) (Hennig, 1966), lost in some avian species: eagle owl, coot and budgerigar, for instance, are devoid of both components; cuckoo, emu and penguin have lost "ε-crystallin"; heron and buzzard do not show the "48 KDa-crystallin", while duck, gull and gannet, like the alligator, still have both compounds in their lenses.

This implies that these components cannot be used as synapomorphies to group avian species together. Yet, a very important lesson may be learnt from these lenticular protein pattern comparisons: figure 6, page 37, clearly shows that in general the same protein bands are found in most lenticular patterns, although relative amounts may differ greatly. This may well imply, that specific protein components being present in low amounts, could escape from detection.

This again would mean that some bands, which seem to be absent, actually are present, but cannot be seen. Assuming that the genetic information for these components has not been lost, it is well possible that their synthesis proceeds at a relatively low rate, depending on the activity of some regulatory control mechanism.

Because of the fact that there exists no principal difference between the presence or absence of some polypeptide chain in the avian lens or the presence or absence of some specific anatomical or morphological feature, it must be concluded that the use of these latter characters as taxonomic instruments may be troubled by the same difficulties. All the perceptible features of an organismal "building plan" are the result of, and under control of, a highly specialized, complicated and integrated unity of regulatory mechanisms. Because of the circumstance that these control systems may bring about some very gradual changes in the morphological-anatomical appearance, transition states may come into existence which escape the human perception. This is the reason that Archaeopteryx may as well be considered a feathered coelurosaur, as a toothed bird (Padien, 1982). Nowadays the distinction between a bird and a reptile is not too difficult, but the recognition of gradual transitions is very troublesome, and may well be highly subjective. This may also hold true for the appreciation of changes occurring in some solitary morphological or anatomical character state, which is used to perform a cladistic analysis. Cladistic analyses, as previously outlined, should result in a direct diagrammic representation of the phylogeny of a given set of taxa, recognition of generally accepted synapomorphies at the anatomical-morphological level, however, may be utterly difficult, due to the possibility of incorrect interpretations of anatomical structures, and convergences. It would, therefore, be most convenient if objective, easily interpretable traits were at hand. Fortunately, it seems that these can be obtained from macromolecular sequences.

II-5: Why macromolecular sequence comparisons?

One of the paradigms of modern biology is the notion that each phenotypical character is determined by genotypical factors which have turned out to be specific nucleotide sequences of the chromosomal DNA. In other words, every biological manifestation is essentially the result of macromolecular sequences. Chromosomal DNA is translated into protein sequences, some of which will act as structural components, while others will become part of regulatory systems, determining the specific outward appearance and functioning of the organism.

Starting from this principle, it should theoretically be possible to define biological species as "the total of their nucleotide sequences". Changes in these sequences which are dramatic enough to give rise to altered sets of proteins and regulatory mechanisms, will finally lead to what we consider to be a "new" species, or, as we might say in a reductionist way, "a new set of sequences", which will no longer interbreed with the "old sequences".

It has recently become possible to determine these nucleotide sequences, which, again theoretically, opens the way to the description of biological species on the basis of their nucleotide sequences, but so far no serious attempts have been made to solve taxonomic problems by direct DNA sequence comparisons.

Yet, they have been used in an indirect way to obtain answers to phylogenetic problems: Sibley and Ahlquist used the technique of DNA-DNA hybridization to determine passeriform interrelationships (Sibley and Ahlquist, 1980; 1982), and to reconstruct the genealogy of the ratites, kiwis and tinamous (Sibley and Ahlquist, 1981). Also restriction fragment analyses of mitochondrial DNA have been applied to avian phylogeny (Kessler and Avise, 1985).

Although these DNA-techniques have already yielded some interesting taxonomic conclusions, also in other areas of vertebrate phylogeny (Brown et al., 1979), they both do not provide the required shared derived characters to perform cladistic analyses; the information they produce is phenetic.

Starting from the idea that by phenetic taxonomic procedures problems are formulated which may be solved by cladistic approaches, it will be clear that in particular the direct comparison of macromolecular sequences, yielding manageable shared derived characters, seems to be a very attractive way to deal with these problems.

From the foregoing it can be concluded that it is often very difficult to judge objectively the validity of some anatomical-morphological character. A decisive method for the phylogenetic classification of biological species on the basis of shared derived characters may therefore possibly be found in comparative nucleotide sequence studies of the chromosomal DNA.

Although the technical possibilities to do this job are already within reach, it was decided to investigate first which information about avian phylogeny could be obtained from the comparative study of some of the translation products, directly reflecting the sequences of this DNA: the amino acid sequences of proteins.

II-6: Protein sequences in avian phylogeny

Protein characteristics have already been applied, as previously outlined, in solving taxonomic problems, in most cases without much conclusive result (Gysels, 1964; Kitto & Wilson, 1966; Sibley & Ahlquist, 1972). This must be attributed to the circumstance that mere protein electrophoretic mobility, which was the character state concerned in these investigations, in most cases gives too little information to reach sound taxonomic conclusions. More clues can be expected from direct amino acid sequence comparisons of homologous proteins. The idea behind this assumption is that after the divergence of two biological species from a common ancestor the independently accumulated changes in the chromosomal DNA of both sister-species may be reflected in the amino acid sequences of homologous proteins. The sequence differences are expected to be more pronounced when the species under comparison are phylogenetically more remote.

An early example of the usefulness of this method is the phylogenetic tree, based on cytochrome c sequences from a wide variety of biological species, which perfectly reflects the generally accepted ideas of the taxonomic interrelation of the biological higher order taxa (Dayhoff, 1972).

Although many more protein sequences have been determined, which have significantly contributed to the understanding of phylogenetic relationships (Goodman, 1982), few attempts have been made to elucidate the many avian evolutionary uncertainties by application of these techniques. Some phylogenetic conclusions, however, were drawn from comparative sequence determinations of cytochrome c (Howard et al., 1974) and lysozyme c (Jolles et al., 1979). Unfortunately, both examples happen to illustrate some of the major pitfalls of comparative amino acid sequence determinations.

Valuable phylogenetic information can be obtained from comparative sequence determination, provided that some important conditions are met:

1. The sequences to be compared, should not only be homologous, but also orthologous (Fitch, 1970). This means that the homology of these proteins should have arisen from speciation, not via gene duplication. Homology originating from gene duplication is properly called paralogy. The evolutionary relationships between orthologous sequences should be congruent with those of the species from which the proteins have been obtained.

Chicken and duck lysozyme c appeared to have similar amino acid sequences, showing extensive immunological crossreaction (Wilson et al., 1977). The lysozyme from goose, however, did not crossreact with either of these two lysozymes, although geese are phylogenetically closely related to the ducks.

The explanation for this odd phenomenon is, that goose lysozyme appears to be paralogous, rather than orthologous, to chicken and duck lysozyme c. It is the product of another genetic locus, arisen by gene duplication, long before the divergence of galliforms and anseriforms, as could be deduced from the circumstance that in black swan both types of lysozyme c were found (Arnheim, 1973).

Misinterpretations of this kind may easily occur when genes, coding for the amino acid sequences to be studied have undergone one or more duplications during their evolutionary history, as is frequently the case.

2. The orthologous genes in both species must have accumulated sufficient mutations to result in informative amino acid replacements.

   Not all nucleotide substitutions in the chromosomal genes are reflected by amino acid replacements in the corresponding polypeptide chains, due to the degeneracy of the genetic code, and if molecular evolution has been too slow, it is very possible that no informative shared derived replacements can be found for the construction of an evolutionary tree.

This is actually the case with avian cytochrome c.

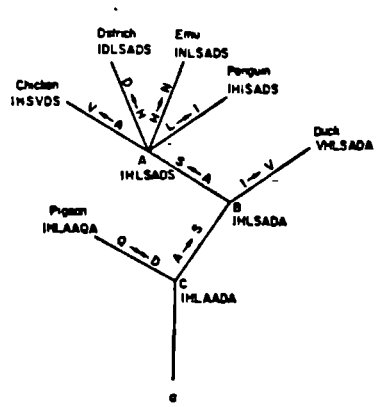


Figure 7: Phylogenetic tree on the basis of cytochrome c sequences. From: Howard et al. (1974).

The constructed tree hardly contains any phylogenetic information, although more information than the authors themselves were able to detect (see: Section III-5).

If molecular evolution on the other hand has proceeded too rapidly, the many occurring amino acid replacements are no longer phylogenetically interpretable.

In other words: it is of major importance to find the right polypeptide chain for sequence comparison, and the tool chosen for the following comparative study is an avian lenticular protein, αA-crystallin.

REFERENCES-II

- Arnheim, N. et al. (1973) J. Biol. Chem. 248: 8036-8042.

- Avise, J.C. & Aquadro, C.F. (1982) Evol. Biol. 15: 151-185.

- Brown, W.M. et al. (1979) Proc. Natl. Acad. Sci. USA 76: 1967-1971.

- Dayhoff, M.O. et al. (1972) "Atlas of Protein Sequence and Structure" Vol. 5, National Biomedical Research Foundation, Washington D.C.

- Fitch, W.M. (1971) Syst. Zool. 19: 99-113.

- Goodman, M. (1982) "Macromolecular Sequences in Systematic and Evolutionary Biology" (Goodman, M., ed.) Plenum, New York.

- Gysels, H. (1964) Thesis. Natuurwet. Tijdschr. 46: 43-178.

- Hennig, W. (1966) "Phylogenetic Systematics" (Davids, D.D. and Zangerl, R. eds.) Univ. of Illinois Press, Urbana.

- Howard, N.L. et al. (1974) Comp. Biochem. Physiol. 48B: 75-85.

- Jolles, J. et al. (1979) Biochemistry 18: 2744-2752.

- Kessler, L.G. and Avise, J.C. (1985) Mol. Biol. Evol. 2 (2): 109-125.

- Kitto, G.B. & Wilson, A.C. (1966) Science 27: 43-178.

- Laemmli, U.K. (1970) Nature 227: 680-685.

- O'Farrell, P.H. (1975) J. Biol. Chem. 250: 4007-4021.

- Padien, K. (1982) Third N. Am. Paleo. Conv. Proc. 2: 387-392.

- Prager, E.M. et al. (1976) J. Molec. Evol. 8: 283-294.

- Sibley, G.C. (1970) Bull. Peabody Mus. Natl. Hist. 32: 1-131.

- Sibley, G.C. & Ahlquist, J. (1972) Bull. Peabody Mus. Natl. Hist. 39: 1-276.

- Sibley, G.C. & Ahlquist, J. (1980) Proc. 17th Int. Ornithol. Congr. Berlin, 1978.

- Sibley, G.C. & Ahlquist, J. (1981) in: "Evolution Today" (Scudder, G.G.E. & Reveal, J.L., eds.) Proc. Sec. Int. Congr. Syst. Evol. Biol.: Hunt Inst. Bot. Doc., Pittsburg, Pa: 301-335.

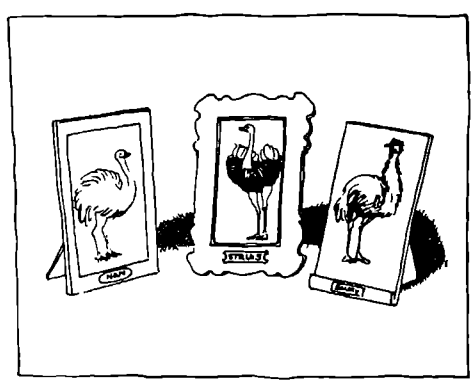
- Sibley, G.C. & Ahlquist, J. (1982) "Postilla" 187: 1-19.

- Weber, K. et al. (1972) Methods Enzymol. 26: 3.

- Wilson, A.C. et al. (1977) Ann. Rev. Biochem. 46: 573-639.

Chapter III

α-Crystallin and Avian Phylogeny



III-1: Why α-crystallin A?

The main reason to choose the αA-chain as a taxonomic tool for the elucidation of avian phylogeny, was the fact that it had previously proven its usefulness in the study of mammalian phylogeny.

It appeared to be a single-gene product (King and Piatigorsky, 1983), having evolved at such an evolutionary rate that sufficient taxonomically informative amino acid replacements had been accumulated among higher order vertebrate taxa.

Cladistic analysis of 42 mammalian αA-chains, including chicken and frog for outgroup comparison, resulted in a phylogenetic tree, based upon shared derived amino acid replacements, which not only reflected the generally accepted biological opinions, but also provided new evidence for the solution of mammalian taxonomic problems (De Jong et al., 1984).

Comparative αA-sequence studies resulted, for instance, in strong evidence for the position of the order Tubulidentata (aardvark) close to the Paenungulata (sea cows, hyraxes and elephants) (De Jong et al., 1981), while other assumed phylogenetic mammalian relationships were confirmed.

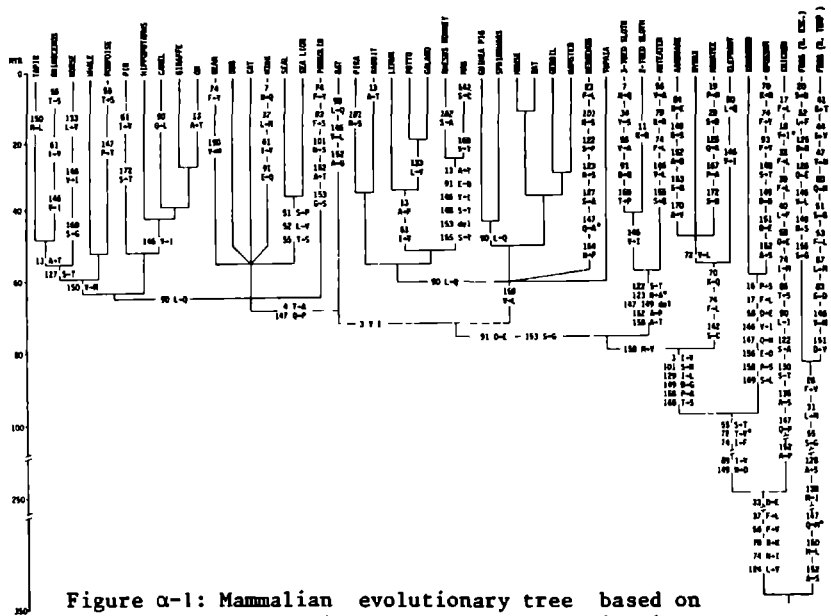


Figure α-1: Mammalian evolutionary tree based on comparative αA-sequence determination. From: De Jong et al. (1984).

Because of the circumstance that the adaptive radiation of mammalian and avian orders is assumed to have proceeded over a comparable span of time (Romer, 1966) and that birds, like mammals, are supposed to possess a single gene in their haploid genome for the αA-chain (Dodemont, 1984), it seemed reasonable to test this polypeptide chain for its phylogenetic information contents on avian orders.

It appeared that sufficient lens material from representatives of different avian orders could be obtained, and that no major problems were to be expected with the determination of avian αA amino acid sequences.

The proposed strategy was the same as in the mammalian study: isolated αA-chains from various species were to be degraded chemically and enzymatically, and the resulting small peptides isolated by peptide mapping.

The amino acid compositions of these peptides were to be compared with the corresponding ones of a reference protein, in this case: chicken αA, of which the amino acid sequence was almost completely known at the time this investigation was to begin. The complete chicken αA-sequence was published some time later (De Jong et al., 1984), and appeared to be in complete accordance with the previously determined chicken αA-sequence, deduced from cDNA-sequencing (Yasuda, personal communication, 1982).

When no differences in amino acid composition between the "unknown" and the "reference" peptides were observed, it was assumed that these similar compositions resulted from similar sequences. This assumption had previously been proven to be acceptable in the case of the comparisons of small peptides (Van Druten et al., 1978).

Peptides showing amino acid compositions differing from the reference peptide were subjected to amino acid sequence determination, using the dansyl-Edman procedure as described by Driessen et al. (1981), to localize amino acid replacements. In this way the complete αA-sequences of turkey, Peking duck, mute swan, crow, eagle owl, oystercatcher, coot, buzzard, gannet, rhea and emu were determined, while the nearly-complete sequences of budgerigar, pigeon, guillemot, black-headed gull, curlew, fulmar, gentoo penguin and ostrich could also be included in the computer analyses used for the assessment of cladistic relationships among the investigated species.

The αA-isolations, sequence determinations, phylogenetic implications and other points of discussion will be outlined in the following sections.

III-2: αA-isolation from Peking duck lenses

If sufficient avian lenses can be obtained, the αA-chain can relatively easily be isolated, a.o. by virtue that it is part of a water-soluble high molecular weight aggregate, the α-crystallin molecule.

This α-crystallin protein, present in the lenses of a wide variety of vertebrate species, is composed of two different types of polypeptide chains, the acidic αA- and basic αB-chains, each about 170 amino acid residues long, being homologous products of a common ancestral gene. These subunits, mostly referred to as αA_2 and αB_2, to discriminate them from their post-translationally modified forms αA_1 and αB_1, occur in varying ratio's, depending on the species studied, in the α-crystallin aggregate, which has in avian lenses an approximate Mr of 600×10^3.

The αA-chain has recently drawn the attention by the circumstance that it has enabled the observation of a unique mRNA-splicing abberation, leading to an elongated αA-chain in rodents (Cohen et al., 1978), and by its sequence homology with the small heatshock proteins of Drosophila (Ingolia and Craig, 1982).

The αA-isolation and sequence determination is exemplified below for Peking duck αA-crystallin. Unless mentioned otherwise, no essential differences exist with αA-isolations and elaborations of other avian species.

1. Eye lenses

Duck lenses were excised from fresh duck heads. It turned out to be very easy to get the lens out of the eye by incision of one eye corner, then clipping the eye open along its edge and softly pressing the eye ball. In this way a hundred duck lenses, weighing 100 mg each, were obtained.

It should be mentioned, however, that it is not obligatory to start with fresh material. It is also possible to isolate proper αA-material from frozen stored bird eyes, or from guanidine hydrochloride stored material (De Jong et al., 1984).

2. αA-Isolation

After homogenization of lens material in two volumes of 1% ammonium bicarbonate pH 7.9, with a Potter-Elvehjem apparatus, centrifugation was performed for 20 min at 15,000 g, resulting in a supernatant containing all major water-soluble lenticular proteins, being the α-crystallin aggregate (composed of αA- and αB-chains), tetrameric δ-crystallin, trimeric ε-crystallin, heterogenous β-crystallin aggregate, and monomeric 48 kDA-crystallin (fig. M-1). This supernatant, containing about 700 A_{280} Optical Density Units, was applied to a column (120 x 3.5 cm) of Ultrogel AcA 34 (LKB), and eluted at room temperature, using the homogenization buffer as eluens, at a flow rate of 30 ml/h.

This procedure resulted in the elution profile of figure M-2. Water-soluble lenticular proteins are thus separated on the basis of their molecular mass. Pooled and lyophilized fractions are subjected to slab gel electrophoresis in the presence of sodium dodecyl sulfate (Laemmli, 1970) to test their identity and purity (fig. M-2-insert). In the case of Peking duck crystallins, also Circular Dichroic spectroscopy of millipore-filtered peak fractions was involved in the characterization of these proteins (see also page 46 for the obtained secondary structure values). The CD-spectra are presented in figure M-3, and revealed that the secondary structures of the crystallins are quite different.

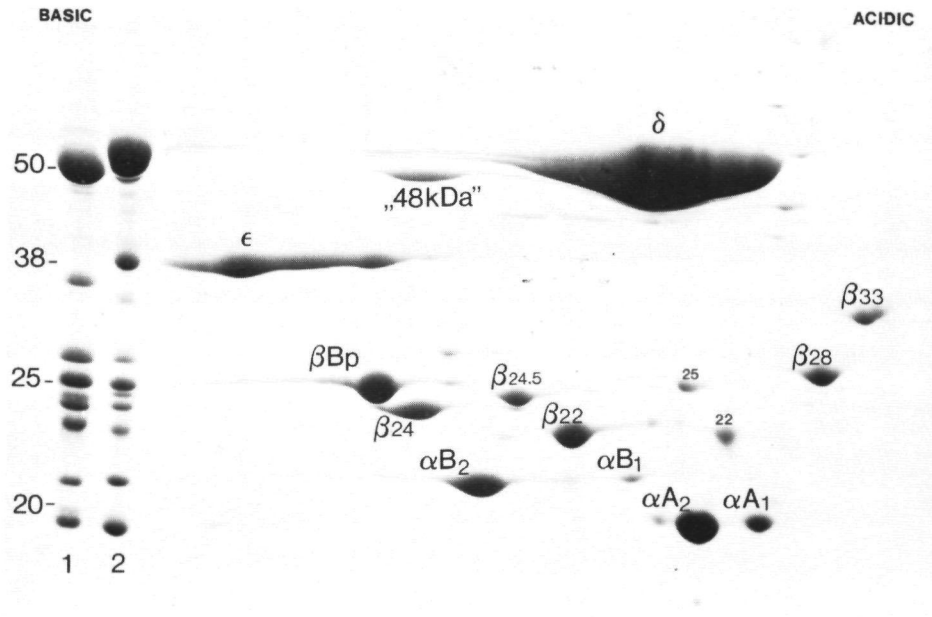


Figure M-1: Two-dimensional electrophoresis of Peking duck lens extract, using 1% ampholine (pH 3.5-10); 1.3% ampholine (pH 7-9) and 2.6% ampholine (pH 6-8) in the focusing gels.

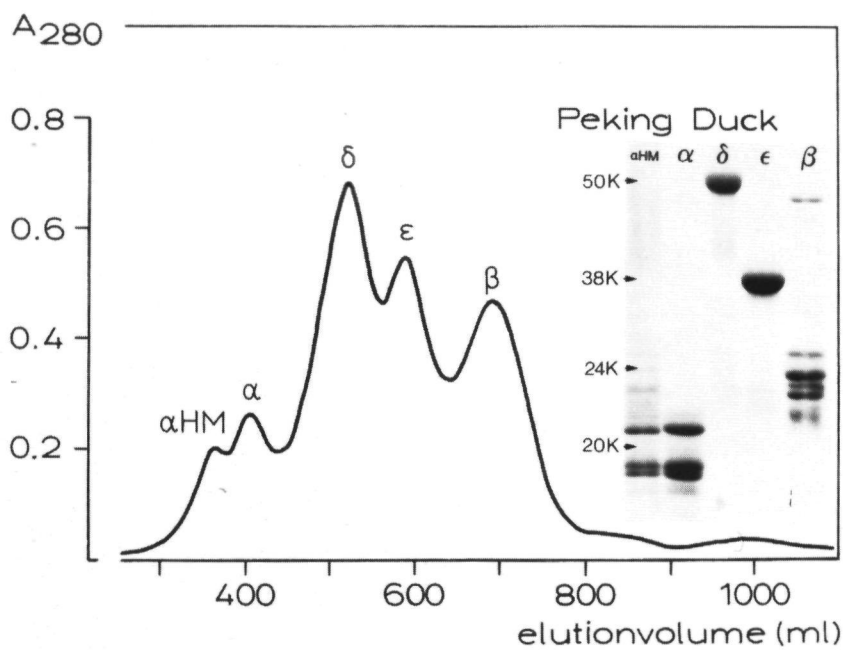


Figure M-2: Elution profile of Peking duck water-soluble lens proteins. Insert: SDS-PAGE of peak fractions.

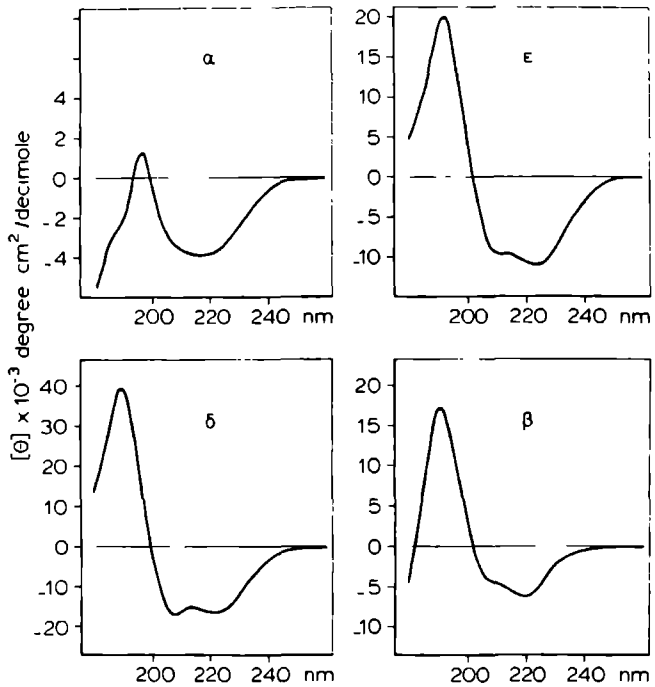


Figure M-3: Circular Dichroism spectra of water-soluble Peking duck lenticular proteins.

After pooling and lyophilization of the Peking duck α-crystallin-containing fractions, about 150 mg of α-crystallin was obtained, which had to be subjected to an ion-exchange chromatographic procedure, using a 0.8 x 22 cm column of carboxymethyl cellulose (Whatman CM-52) and a linear gradient of 0.4-0.12 M sodium acetate, pH 5, in 8 M urea, to isolate the αA-chain (fig. M-4).

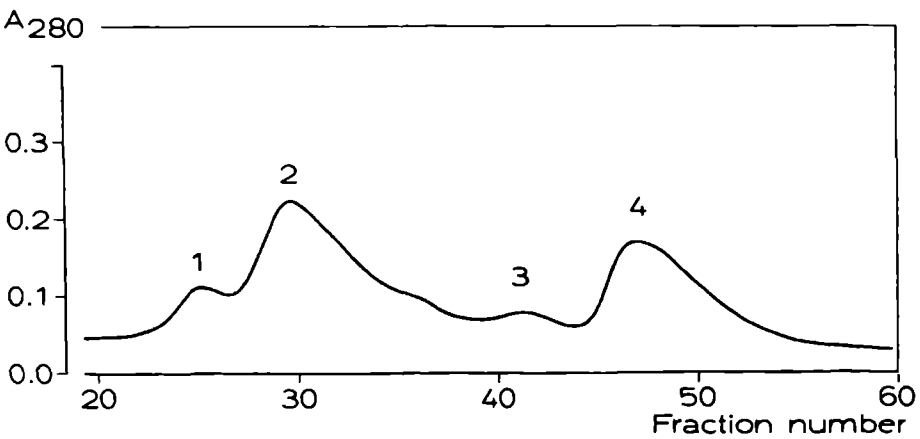


Figure M-4: Ion-exchange chromatographic profile of Peking duck α-crystallin.

Pool 2 contains the primary gene product αA_2, while pool 1 represents the post-translational charge-modification αA_1. The same goes for pools 4 and 3, containing αB_2 and αB_1, respectively.

The SDS-electrophoretic patterns and alkaline-urea polyacrylamide gel electrophoretic patterns of these pools are shown in figures M-5 and M-6, indicating that pool 2 contains a relatively pure αA-fraction which is ready to be used for further elaboration, in order to elucidate its amino acid sequence.

It turned out that about 50 mg of pure αA-protein could be isolated from 100 duck lenses.

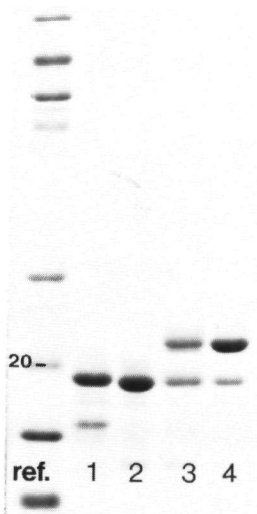
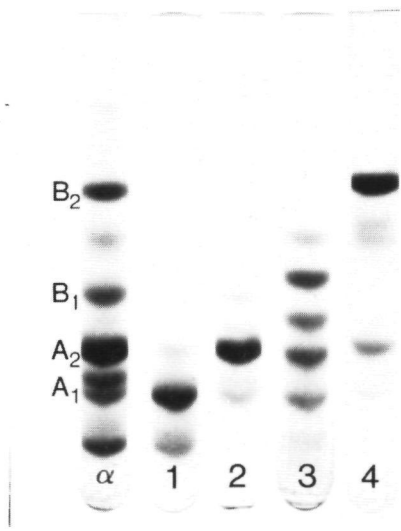


Figure M-5: SDS gel electrophoretic patterns of ion-exchange pool 1/4. Left: MW marker protein.

Figure M-6: Alkaline-urea gel electrophoretic patterns of ion-exchange pool 1/4. α: total α-crystallin.



III-3: αA-sequence determination

To deduce the amino acid sequence of duck αA-crystallin, comparisons of amino acid compositions of small duck peptides and chicken reference peptides were made, as stated before.

To obtain these small peptides, the αA-chain was subjected to S-β-aminoethylation, resulting in an additional trypsin-sensitive site by chemical modification of the cysteinyl residue at position 131. Then a small portion of the total material (about 4 mg), was digested with trypsin.

The rest is chemically fragmented by treatment with cyanogen bromide (CNBr), as described by Driessen (1981). This results in a mixture of four fragments, which is applied to a gel filtration column of Sephadex G50 sf in 5% acetic acid. The elution pattern is shown in figure M-7.

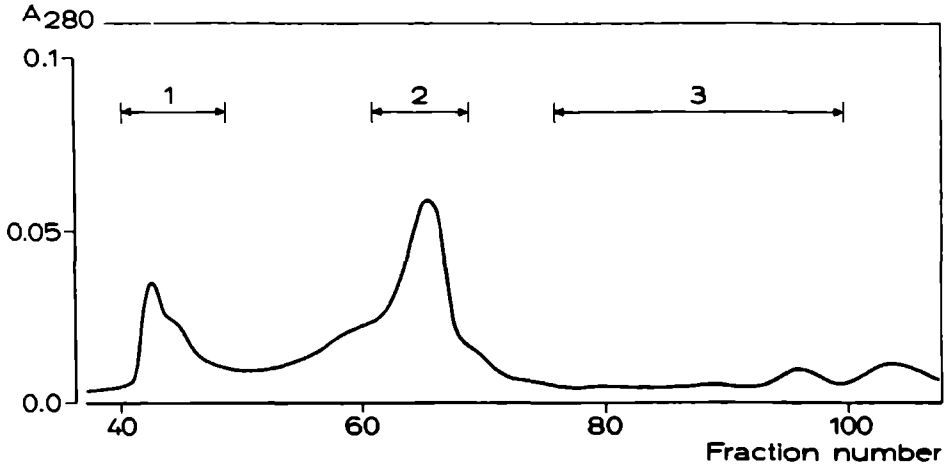


Figure M-7: Gel filtration of CNBr-cleaved Peking duck αA.

Pool 1 contains uncleaved material, while peak 2 represents the two major CNBr-fragments: res. 2-74 and res. 75-138. Pool 3 contains two smaller fragments: res. 139-150 and res. 151-173. Methionyl-residue 1 is split off and lost in this procedure, which is the reason that some of the protein material is not subjected to CNBr-treatment, but completely digested with trypsin, which also gives the opportunity to designate some overlap peptides.

After lyophilization of pools 2 and 3, tryptic digestion is carried out, followed by separation of the tryptic peptides by high voltage paper electrophoresis at pH 6.5 and descending chromatography as described elsewhere (De Jong et al., 1981). Peptide maps were stained with Ninhydrin or Fluorescamine (Fluram, Roche) (Udenfreund et al., 1972), to enable the detection of the peptides for amino acids analyses.

This strategy, based on the CNBr-treatment, turned out to be a very useful one because of the fact that so far it never had been possible to find the tryptic peptide res. 71-78. CNBr-treatment, however, splits this peptide into two easily detectable smaller peptides by the circumstance that res. 74 appeared to be a methionyl residue. This finding made it possible to complete

the chicken αA-sequence, functioning as a reference-sequence for the other unknown avian αA-chains.

Although the main part of the αA-chain could be mapped in this way, another part remained "out of sight". This missing information is partly represented by a number of peptides which can be found, after electrophoresis and chromatography, in the so-called "neutral zone" (NZ), which harbors the peptides uncharged at pH 6.5. These peptides can be separated by a second electrophoretic step. The "NZ" is cut out of the original peptide map, applied on a new strip of Whatman-paper and subjected to electrophoresis at pH 3.5.

The last problem is the large tryptic core-peptide res. 22-49, which precipitates after trypsin treatment and is insoluble at pH 6.5. This last "hidden part" of the αA-chain can be included in the peptide comparisons by washing this insoluble precipitate three times with 0.01 M HCl, followed by thermolytic digestion. This results in 7 small peptides which can easily be separated by peptide mapping.

By application of the complete procedure, all the peptides of the αA-chain can be found and compared to the reference sequence. The following table (table M-1), summarizes the results of peptide-detection after cyanogen bromide cleavage of the avian αA-chain, while figure M-8A,B,C, indicates the map positions of the Peking duck αA tryptic and thermolytic peptides. Figure M-9 shows the complete Peking duck α-crystallin A sequence, with threonine at position 153, instead of the chicken serine-153.

Table M-1: Tryptic peptides obtained by trypsin treatment of CNBr-pools 2 and 3 (fig. M-6).

CNBr-pool	Tryptic peptide	Residues
	CB-T1*	2-11
2	T2	12
	T3	13-21
	T4 (insoluble core-peptide)**	22-49
	T5	50-54
	T6	55-65
	T7 (neutral-zone peptide)	66-68
	T8 (neutral-zone peptide)	69-70
	T9a	71-74
	T9b (neutral-zone peptide)	75-78
	T10	79-88
	T11	89-99
	T12	100-103
	T13 (neutral-zone peptide)	104-112
	T14	113-116
	T15	117
	T16	118-119
	T17a (neutral-zone peptide)	120-131
	$T17b_1$	132-138
	$T17b_2$	139-145
3	T18-19a	146-150
	T18-19b	151-163
	T20	164-173

* CB-T1 lacks the methionyl-residue 1, which is found by tryptic digestion of non CNBr-cleaved αA-material, yielding T1(res1-11).

** Thermolytic subdigestion of T4 results in seven thermolytic (Th-) peptides: Th1, res. 22-26; Th2, res. 27-30; Th3, res. 31-35; Th4, res. 36-39; Th5, res. 40-43; Th6, res. 44-47 and Th7, res. 48-49.

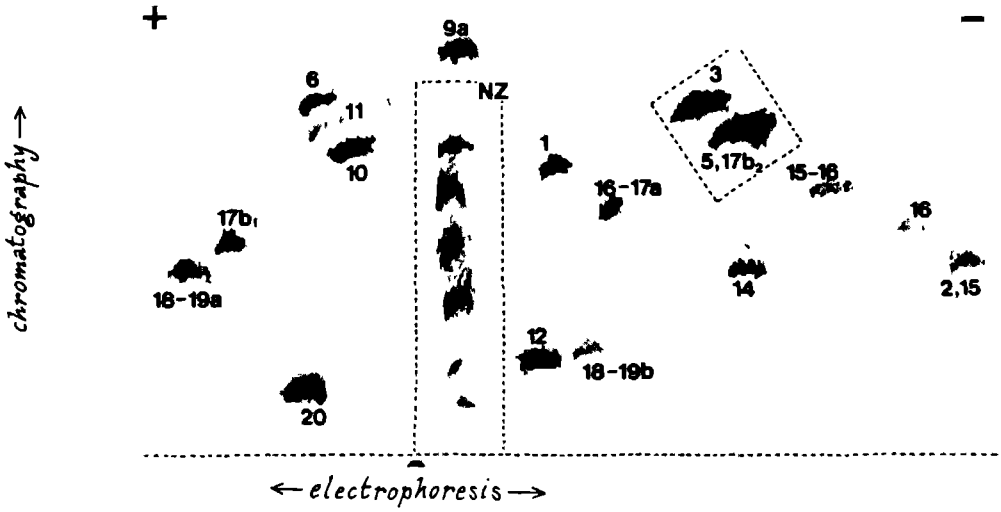


Figure M-8A: Tryptic peptides of CNBr-pool 2 (Table M-1).

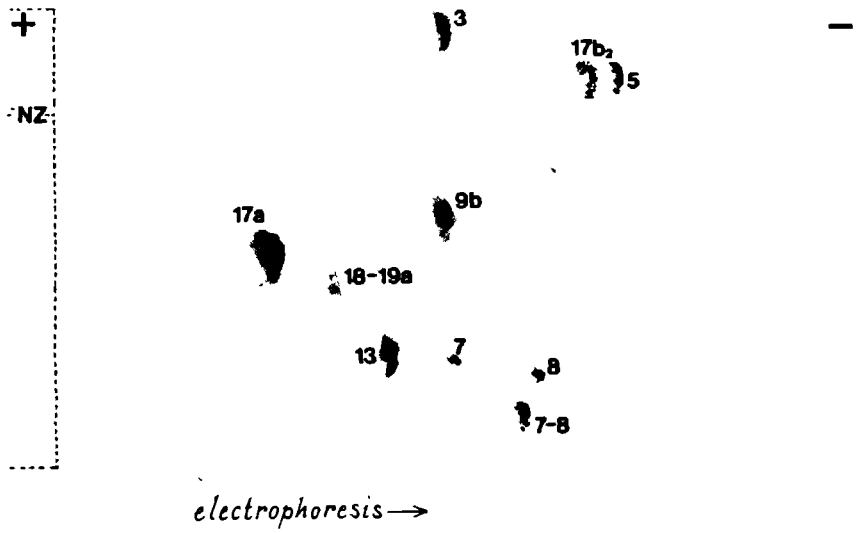


Figure M-8B: pH 3.5 reelectrophoretic separation of T5 and T17b$_2$ and "Neutral Zone"-peptides.

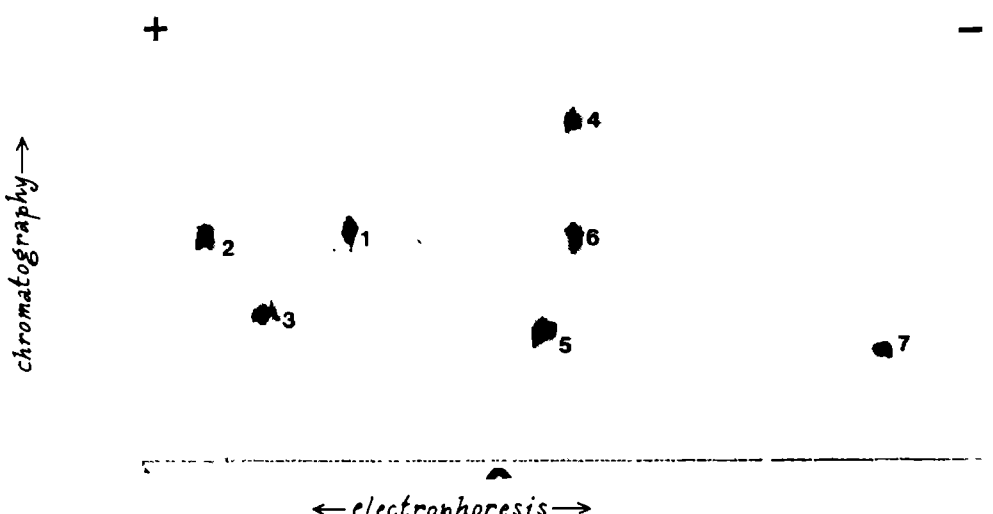


Figure M-8C: Thermolytic peptides from tryptic core-peptide 22-49.

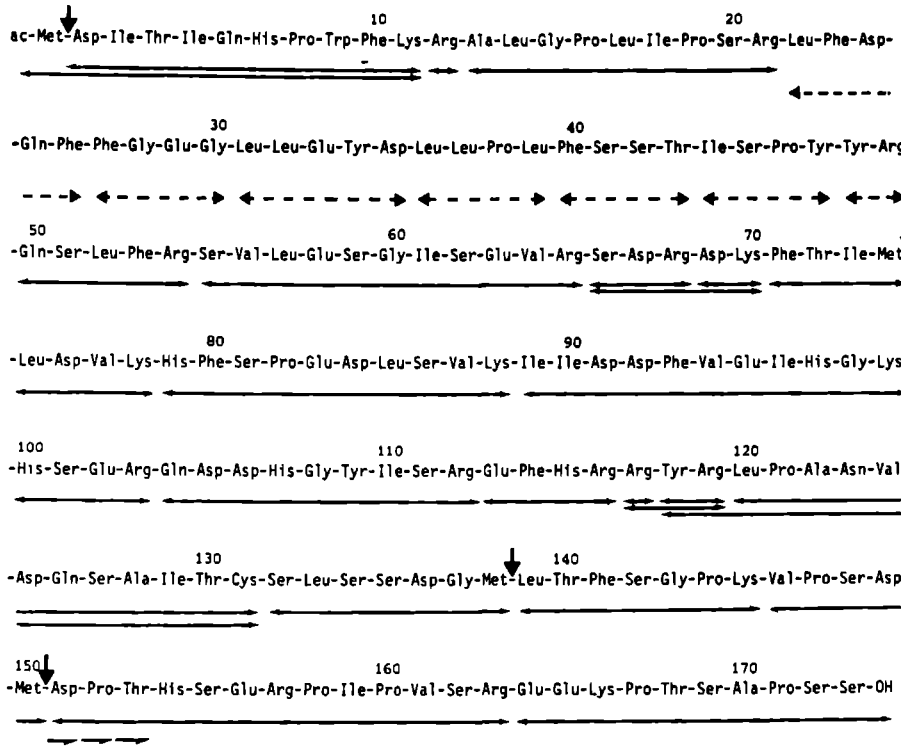


Figure M-9: Peking duck α-crystallin A amino acid sequence. (⟷) tryptic and (⟷) thermolytic peptides. Arrows (⬇) indicate CNBr-cleavage points. (⟶⟶) indicate dansyl-Edman degradation steps to localize amino acid replacements.

After some time, however, it appeared that the time-consuming gel-filtrational separation of CNBr-fragments could be omitted. This finding reduced the amino acid sequence determination of an unknown avian αA-chain to:

1. S-β-aminoethylation and cyanogen bromide cleavage

2. tryptic digestion of the total CNBr-cleaved material

3. washing the remaining precipitate with 0.01 M HCl

4. thermolytic digestion of this precipitate

5. peptide mapping of tryptic and thermolytic peptides

6. pH 3.5 reelectrophoretic separation of overlapping peptides ("Neutral Zone"-peptides and T5/T17b_2)

7. amino acid composition comparisons to reference chicken peptides

8. dansyl-Edman sequence determination to localize amino acid replacements

This procedure, which was in most cases performed succesfully, indicates that the different avian water-soluble lens proteins are present in varying relative amounts, as can be concluded from the gel filtration patterns of Peking duck, curlew, eagle owl and emu (fig. M-10).

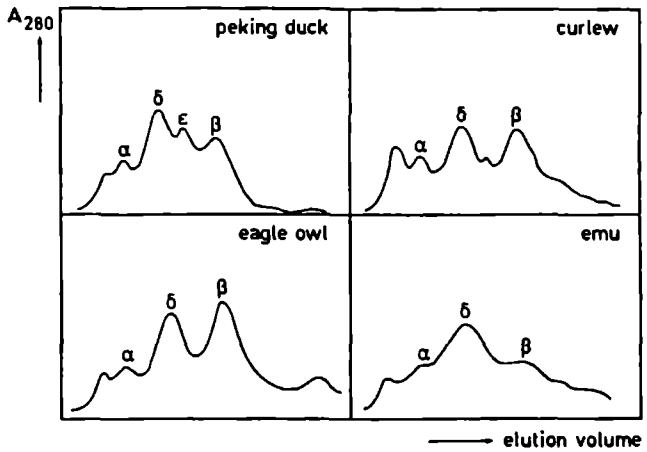


Figure M-10: Gel filtration profiles of water-soluble lenticular proteins from four different avian species.

After checking the purity of individual protein pools by SDS-gel electrophoresis, it appeared that in a number of cases α-crystallin pools were contaminated with other crystallins. Particularly in the emu gel filtration β-crystallin components could be found all over the pattern. In those cases, α-crystallin enriched fractions were pooled, concentrated by Amicon filtration, to avoid possible undesired lyophilization artefacts, and reapplied onto the same column, which in all cases resulted in an improved purity of the α-crystallin pool.

In some cases, the amount of α-crystallin did not seem to be sufficient to perform ion-exchange chromatographic isolation of the αA-chain, because of the possible losses of protein material resulting from this procedure. Because of the circumstance that the relative amount of αA-chain in the α-crystallin aggregate exceeds the amount of αB, it was in a number of cases (curlew, guillemot, oystercatcher, fulmar and budgerigar) also possible to omit the ion-exchange step and to use complete α-crystallin for the peptide comparisons.

A special case, however, was represented by the αA-isolation from ostrich lenses, which were transported in guanidine hydrochloride from South Africa. It appeared not to be possible to reassociate the crystallin polypeptides into their original aggregates (see Bloemendal et al., 1975), so the αA-chain had to be directly purified from the total mixture of crystallin subunits by ion-exchange chromatography.

By application of this technique an αA-enriched fraction was isolated, but it was not possible to trace CB-tryptic peptide 1 (T1, res. 2-11), which, however, had no consequences for the phylogenetic conclusions.

Also in some other cases there was not sufficient α-crystallin material available to complete the amino acid sequence of αA, and sometimes no proper integral values could be obtained with the amino acid analyses of some peptides. In those cases, these values were not included in the computer analyses for the construction of an evolutionary tree, leaving the αA-chains of a number of birds incomplete.

For these reasons, a number of peptides is missing: budgerigar T10; pigeon T17a; guillemot, gull and heron T9a, curlew T5,6,7,8,10,11,17b_1; fulmar T6,7,8,9a,11 and 18-19a; penguin T17b_1 and ostrich t1, as mentioned before.

In most cases, however, incomplete chains were taken into account for the construction of a cladogram, which is presented in the following section.

III-5

A cladistic analysis of 21 avian αA-chains

Nature 311, 5983-5985 (1984)

Reprinted from Nature, Vol 311, No 5983, pp 257-259, 20 September 1984

Ratites as oldest offshoot of avian stem—evidence from α-crystallin A sequences

Steven O. Stapel*, Jack A. M. Leunissen*, Marlies Versteeg*, Jan Wattel† & Wilfried W. de Jong*

* Department of Biochemistry, University of Nijmegen, Geert Grooteplein Noord 21, 6525 EZ Nijmegen, The Netherlands
† Institute of Taxonomic Zoology, University of Amsterdam, 1018 DC Amsterdam, The Netherlands

One of the most disputed issues in avian phylogeny is the origin of the ratites, the large flightless birds of the Southern Hemisphere (reviewed in refs 1–3). It is still not generally agreed whether the ostriches, rheas, emus and cassowaries, and probably kiwis, form a natural, monophyletic group, although much recent evidence supports this view[4–6]. Also, their phylogenetic relationship with the other avian orders remains unresolved; comparative protein sequence studies might shed new light on this problem. Therefore, we determined the amino acid sequence of the eye lens protein α-crystallin A in ostrich, rhea and emu, and in representatives of 13 other avian orders. Comparison of these sequences with known αA sequences of mammals, reptiles, frog and dogfish provides strong evidence that the ratites, as a monophyletic assemblage, represent the first offshoot of the avian line.

The α-crystallin A-chain is a suitable protein for comparative sequence analysis because (1) being a major eye lens constituent in all vertebrate classes[7], it can often be isolated in considerable quantities, and (2) the sequence analysis of the 173-residues αA chain is relatively simple[8] There is only a single gene for αA[9], which excludes the possibility of errors resulting from comparing paralogous[10] products of multi-copy genes in different species The rate of evolution of αA is sufficiently slow to be informative in the study of more distantly related taxa[11] Sequences of αA-chains have been determined previously in 43 mammalian species[8,9], chicken[8], alligator (W W de J, M V, A Zweers, H Dessauer and M Goodman, unpublished data), and tegu[11], frog[8,12] and dogfish[11] These studies have already provided useful information about some aspects of mammalian phylogeny[11–13], and allow ample outgroup comparisons for the study of avian phylogeny In the present study we have included, apart from the large ratites, several other avian orders for which the phylogenetic relationships are disputed To confirm that closely related species indeed have related αA sequences, which would support the reliability of our approach, we included in our study chicken, turkey (both Galliformes), domestic duck and mute swan (Anseriformes) We obtained sufficient eye lenses to perform sequence analysis of the αA-chains from 21 avian species representing 16 orders (Table 1)

The methods of isolation and sequence analysis of the avian αA-chains and the resulting sequence information, as exemplified by the αA-chain of the rhea, are given in Fig 1 and its legend The sequence differences between the avian αA-chains are shown in Table 1, relevant sequence data from outgroup species are included The six mammalian species represent the major lineages of the mammals investigated previously[11] Table 1 immediately reveals the common occurrence in the ratite species of the unique substitution, Asp 69→Glu Another character shared by the ratites is residue Ala 148, which, however, also occurs in pigeon, gull and guillemot Because Asp 69 and Ser 148 seem to be the primitive character states present in nearly all outgroups, these substitutions may be considered as shared derived (synapomorphous) characters, in support of a ratite monophyly Equally conspicuous is the presence of Ala 122 and Pro 147 as apparently shared derived characters in all birds, apart from the ratites Residues Ala 127 (or Thr) and Asn 135 again seem to be synapomorphies of all birds, with the exception of ratites, galliform and anseriform birds

It is obvious that parallel and back-substitutions do occur in the evolution of αA, as in all other proteins This complicates the interpretation of differences between homologous amino acid sequences, and necessitates the use of computer analyses to objectively assess the cladistic relationships among the investigated sequences The objectives of such computer searches are to find the most parsimonious phylogenetic trees (that is, branching patterns that require the fewest amino acid substitutions) and to calculate the cost, in terms of additional required substitutions, of alternative topologies By applying different

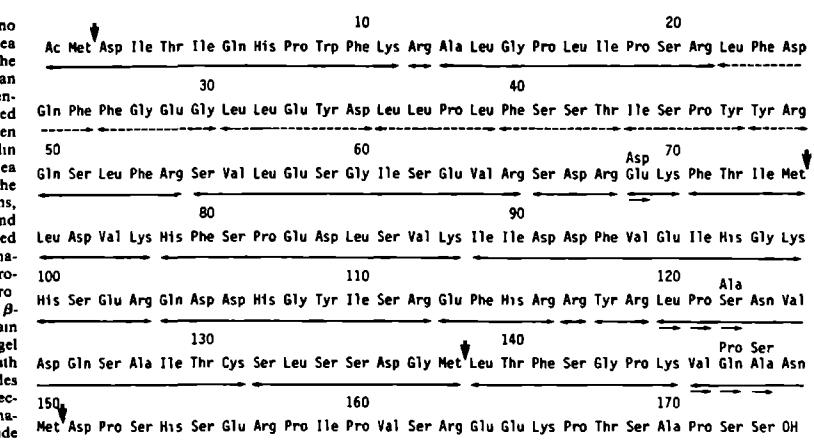


Fig. 1 Proposed amino acid sequence of the rhea α crystallin A chain The characterization of avian αA chains followed the general procedures described for mammalian and chicken αA chains[8] α-Crystallin was isolated from four rhea lenses by gel filtration of the water soluble lens proteins, and the component αA and αB chains were separated by ion exchange chromatography Fragments produced by cyanogen bromide cleavage of the *S*β-aminoethylated αA chain were separated by gel filtration and digested with trypsin Soluble peptides were isolated by paper electrophoresis and chromatography The core peptide (22–49) was purified by gel filtration, then subdigested with thermolysin for peptide mapping Peptide 1–11 was obtained by tryptic digestion of the total αA chain Amino acid compositions of tryptic (↔) and thermolytic (←– – –→) peptides were compared with the corresponding ones of the chicken αA chain[8] When no difference in composition was observed, we assumed identity between the chicken sequences and rhea peptide Peptides which differed in composition from the corresponding chicken peptides were subjected to dansyl–Edman degradation (→) to localize substitutions Residues which are different in the chicken αA chain are shown above the rhea αA sequence Cleavage points of cyanogen bromide are indicated by ↓ The avian αA chains are *N*-terminally blocked, probably by an N^{α} acetyl group as in bovine αA[7]

Table 1 Phylogenetically informative positions in avian, reptilian and outgroup α-crystallin A sequences

```
                                      1111111111111111111111111111111
POSITION NR        1111133334555567778889990122222223333444444455555556667
                 341367823790056890241690161123456790258234678902345682892

OX               IAKTPFYFELFLQTVDDKVFSTVQEINSSNVDQSLSSAMSGPIPSGVAGHSEASSSS
DOG              ||A||||||||||||||||||L||||||||||||||||V||||||||||||
RAT              VT|A|||||||||||||||||L|||||||||||||||VQ| L||||||||
HYRAX            VT|A||||||||||QLL||LD|||||||||||||C||VQ||M|S|||||
2-TOED SLOTH     VTRA|||||||||||||||LD||TA||||||||||||||||PS||T||
KANGAROO         |T|ASL|||||||E|||||LD|S||||I||||||||M|DM|S||DS|TL
CROW             T|A|LIL||LF|S|E||TM|SIID|S|A|||AIT|N|||V||NM||||P|T|
EAGLE OWL        T|A|LIL||LF|S|E||TM|SIID|S|A|||AIT|N|||V||NM||||P|T
BUDGERIGAR       T|A|LIL||LF|S|E||TM  IID|S|A|||AIT|N|||V||NM|S||P|T
PIGEON           T|A|LIL||LF|S|E||TM|SIID|S         N|||V|ANM|||P|T
GUILLEMOT        T|A||IL||LF|S|E|  MNSIID|S|A|||AIT|N||H|V|ANMVS||P|T
GULL             T|A|-I|||LF|S|E|   M|SIID|S|A|||AIT|N|||V|ANM|S||P|T
CURLEW           T|A|LI|||LF        TM      S A|||AIT    V||NM|S||P|T
OYSTERCATCHER    T|A|LIL||LF|S|E||TM|SIID|S|A|||AIT|N|||V||NM|S||P|T
COOT             T|A|LIL||LF|S|E||TM|SIID|S|A|||AIT|N|||V||NM|S||P|T
BUZZARD          T|A|LI|||LF|S|E||TM|SIID|S|A|||AIT|N|||V||NM|||P|T
HERON            T|A|LIL||LF|S|E||  M|SIID|S|A|||TIT|N|||V||NM|S||P|T
GANNET           T|A|LIL||LF|S|E||TM|SIID|S|A|||AIT|N|||V||NM|||P|T
FULMAR           T|A|LIL||LF          S      S|A|||AIT|N        M|S|T|P|T
PENGUIN          T|A|LIL||LF|S|E||TM|SIID|S|A|||TVT    V|V||NM|SR||P|T
DUCK, SWAN       T|A|LIL||LF|S|E||TM|SIID|S|A||||IT|S|||V||NMPT||P|T
CHICKEN TURKEY   T|A|LIL||LF|S|E||TM|SIID|S|A||||IT|S|||V||NMPS||P|T
►OSTRICH           A|LIL||LF|S|EE|TM|SIID||||||IT|S|||VQANMPS||P|T
►RHEA            T|A|LIL||LF|S|EE|TM|SIID|S||||IT|S|||VQANMPS||P|T
►EMU             T|A|LIL||LF|S|EE|TM|SIID|SA|||IT|S|||VQANMPS||P|T
ALLIGATOR        T|A|LI|||L|HS|E||TM|SIIDV|||S|||VTV|||SVQ|N|TI||DPAT
TEGU             T|A|LI|||LF|||E||T||S|I|||||||AIT|||AA|VQ|NTPS|N|P|T|P
FROG             |T|A||||DF|||GF||R|||ILD|S|||LNE|I|||I|||LM|SLSS|G|P|T||
DOGFISH          L|RSS||||DF|P|NF||RMN|SIVD|A|||LNE|IA|NLCC|TRP|DSNWQDP|GTE
```

Positions shown are those which are variable in the αA chains of at least two of the species, plus the autapomorphous substitutions in the avian and reptilian sequences Vertical lines indicate that residues are identical to the topmost sequence Amino acid sequences have been analysed, as outlined in Fig 1 legends, for carrion crow (*Corvus corone*, Passeriformes), eagle owl (*Bubo bubo*, Strigiformes), budgerigar (*Melopsittacus undulatus*, Psittaciformes), domestic pigeon (*Columba livia*, Columbiformes), guillemot (*Uria aalge*, Charadriiformes), black-headed gull (*Larus ridibundus*, Charadriiformes), curlew (*Numenius arquata*, Charadriiformes), oystercatcher (*Haematopus ostralegus*, Charadriiformes), coot (*Fulica atra*, Gruiformes), buzzard (*Buteo buteo*, Falconiformes), heron (*Ardea cinerea*, Ciconiiformes), gannet (*Sula bassana*, Pelecaniformes), fulmar (*Fulmarus glacialis*, Procellariiformes), gentoo penguin (*Pygoscelis papua*, Sphenisciformes), domestic duck (*Anas platyrhynchos*, Anseriformes), mute swan (*Cygnus olor*, Anseriformes), turkey (*Meleagris gallopavo*, Galliformes), ostrich (*Struthio camelus*, Struthioniformes), rhea (*Rhea americana*, Rhei formes), and emu (*Dromaius novae-hollandiae*, Casuariiformes) The αA sequences and scientific names of the mammals, chicken, tegu, frog and dogfish have been given previously[8,11] The alligator is *Alligator mississippiensis* The sequences of chicken and turkey αA were found to be identical, as are those of duck and swan Some αA sequences have been incompletely determined ( ) Thr 127 in heron αA has not been ascertained by Edman degradation, and is therefore not taken as a synapomorphy with penguin αA In sloth αA there is a deletion of three residues (–) The one letter notation for amino acids is A, Ala, C, Cys, D, Asp, E, Glu, F, Phe, G, Gly, H, His, I, Ile, K, Lys, L, Leu, M, Met, N, Asn, P, Pro, Q, Gln, R, Arg, S, Ser, T, Thr, V, Val, Y, Tyr

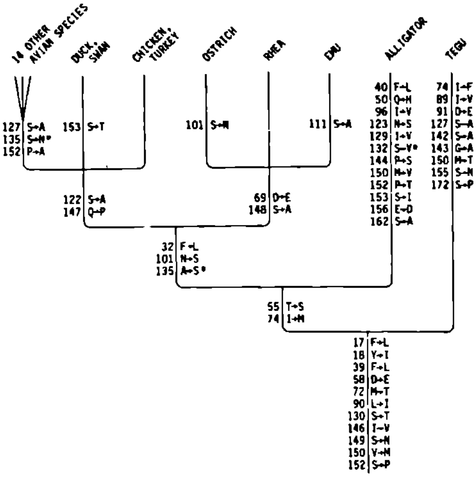


Fig. 2 Avian and reptilian branches of the maximum parsimony tree constructed on the basis of the αA sequence data in Table 1 The procedure starts with the calculation of a matrix of minimum mutation distances for the αA sequences[14], followed by a tree construction programme which begins with an unrooted three-sequence tree to which the other sequences are added one at a time, using a branch-swapping procedure, until a tree of minimal length is found (essentially according to Fitch and Margoliash[14], as modified by J Felsenstein) The tree construction procedure was repeated several times using different input orders of the αA sequences, this resulted in the same topology in all instances, apart from equally parsimonious minor alternatives in the cluster of 14 'other avian species' The number of amino acid substitutions required for the most parsimonious tree was 163 The outgroup part of this tree was congruent with the previously constructed topology[11], and is therefore not shown here Amino acid substitutions in the different branches are deduced from the reconstructed ancestral αA sequences at the nodes[15] Substitutions that could not be assigned unambiguously to a certain branch were, as far as possible, placed on terminal branches, rather than between ancestral nodes[16], this minimizes the synapomorphous substitutions characterizing monophyletic groupings in this cladogram Taking six other mammalian αA sequences (horse, mink, human, elephant, three toed sloth and opossum)[11] as outgroups in the tree construction procedure resulted in the same topology as shown here * Substitutions 132 S→V and 135 A→S (or 135 S→N) require two nucleotide replacements

algorithms[14–16] to the data set of Table 1, we found the most parsimonious tree (see Fig 2), which requires 163 amino acid substitutions All changes in the topology presented in Fig 2 require additional amino acid substitutions (Table 2) On this basis we must conclude that the presence of Ala 148 in ratites and in pigeon, gull and guillemot is due to parallel, or convergent, evolution

The constructed tree clearly supports the hypothesis that the ratites are monophyletic, and depicts them as the sister group of all other birds The Galliformes and Anseriformes appear to be the next to branch off from the main avian stem Our data allow the possibility that Galliformes and Anseriformes are monophyletic[3,4], but conflict with a close relationship between Anseriformes and Charadriiformes[17] (Table 2) The other investigated avian orders are grouped as a monophyletic assemblage, but the scarcity of substitutions in their αA sequences prevents a further, meaningful resolution of their relationships In fact, several species from different orders are found to have identical symplesiomorphous αA sequences The most parsimonious tree of αA also stresses the monophyletic origin of the investigated birds and reptiles, and places the crocodiles, represented by the alligator, as a sister group of the birds These findings conform to the generally held, but not unchallenged, palaeontological views The recently proposed sister group relationship of birds and mammals[18] would, in the case of αA sequences, result in a considerable loss of parsimony (Table 2)

Also, recent immunological[4] and DNA hybridization studies[6] have provided evidence for a monophyletic origin of the ratites, thus contradicting the suggestion that they are paraphyletic or polyphyletic[2,19] In fact, Prager and Wilson[20] have suggested a phylogenetic tree, on the basis of immunological comparisons, in which the position of ratites, Galliformes and Anseriformes relative to the other avian orders is essentially as in our *αA* tree The primitive, rather than derived, position of the ratites within birds is, among others, also supported by karyological[5] and

Table 2 Numbers of additional amino acid substitutions required to change the topology of Fig. 2

Alternative relationships	Additional substitutions
Polyphyletic origin of ratites from the avian stem	≥4
Ratites within the cluster of 'other avian species'	≥5
Ratites and Galliformes (or Anseriformes) monophyletic	2
Galliformes and Anseriformes monophyletic[3]	0
Anseriformes with Charadriiformes[17]	3
Birds and mammals as sister groups[18]	6

ontogenetic[21] studies. However, a phyletically old origin of the ratites could not be concluded on the basis of morphological and anatomical data[1-3]. Comparative sequence data of avian proteins, including one or more ratite species, are available for cytochrome c[22] and haemoglobin[23]. On the basis of the very few substitutions in seven avian cytochrome c sequences, a cladogram has been proposed in which the ratites are depicted as derived from carinate birds[22]. We found, however, that a more parsimonious tree can be obtained from this cytochrome c data set by assuming the ratites to be a sister group of the other birds. The sequence determinations of avian α and β haemoglobin have not yet led to specific statements about ratite phylogeny[23].

Our αA sequence data may have some bearing on the problem of the time of radiation of the avian orders. Estimates for the time since the last common ancestor of all modern birds range between 130 and 65 Myr[6,24]. Considering that αA evolves at an average rate of three amino acid substitutions per 100 residues in 100 Myr[8], it seems that the small numbers of substitutions in the avian αA chains since their divergence from the last common ancestor (Fig. 2) favour the more recent time of avian radiation. Finally, the branching order in our tree may lend support to the idea that sustained flight developed late in avian evolution, after the divergence of ratites and Galliformes from the main stem.

We thank Mr W. B. Appel and the Klein Karoo Landboukoöperasie for ostrich lenses, Dr S. J. Davies (CSIRO) for emu lenses, Drs Th. Smit and Ms T. Prins for other avian lenses, and Dr J. Felsenstein for supplying his Phylip package. This work was supported by the Foundation for Fundamental Biological Research (BION) with financial aid from the Netherlands Organization for Pure Research (ZWO).

Received 10 April, accepted 17 July 1984

1 Sibley, G. S. & Ahlquist, J. E. *Bull. Peabody Mus. nat. Hist.* **39**, 44–52 (1972)
2 Feduccia, A. *The Age of Birds* (Harvard University Press 1980)
3 Cracraft, J. *Auk* **98**, 681–714 (1981)
4 Prager, E. M., Wilson, A. C., Osuga, D. T. & Feeney, R. E. *J. molec. Evol.* **8**, 283–294 (1976)
5 de Boer, L. E. M. *Nature* **287**, 84–85 (1980)
6 Sibley, G. S. & Ahlquist, J. E. in *Evolution Today* (eds Scudder, G. G. E. & Reveal, J. L.) 301–335 (Carnegie Mellon Pittsburgh, 1981)
7 Bloemendal, H. (ed.) *Molecular and Cellular Biology of the Eye Lens* (Wiley, New York, 1981)
8 de Jong, W. W., Zweers, A., Versteeg, M. & Nuy Terwindt, E. C. *Eur. J. Biochem.* **141**, 131–140 (1984)
9 King, C. E. & Patagonsky, J. *Cell* **32**, 707–712 (1983)
10 Fitch, W. M. *Syst. Zool.* **19**, 99–113 (1970)
11 de Jong, W. W. in *Macromolecular Sequences in Systematic and Evolutionary Biology* (ed. Goodman, M.) 75–114 (Plenum, New York, 1982)
12 Tomarev, S. I. *et al. FEBS Lett.* **162**, 47–51 (1983)
13 de Jong, W. W., Zweers, A. & Goodman, M. *Nature* **292**, 538–540 (1981)
14 Fitch, W. M. & Margoliash, E. *Science* **155**, 279–284 (1967)
15 Dayhoff, M. O., Park, C. M. & McLaughlin, P. J. in *Atlas of Protein Sequence and Structure* Vol. 5 (ed. Dayhoff, M. O.) 7–16 (National Biomedical Research Foundation, Silver Springs, Maryland, 1972)
16 Goodman, M., Moore, G. W., Barnabas, J. & Matsuda, G. *J. molec. Evol.* **3**, 1–48 (1974)
17 Olson, S. L. & Feduccia, A. *Smithson. Contr. Zool.* **323**, 1–24 (1980)
18 Gardiner, B. G. *Zool. J. Linn. Soc.* **74**, 207–232 (1982)
19 Houde, P. & Olsen, S. L. *Science* **214**, 1236–1237 (1981)
20 Prager, E. M. & Wilson, A. C. in *Acta 17th int. Orn. Congr.* (ed. Nöhring, R.) 1209–1214 (Deutsche Ornithologen, Berlin, 1980)
21 McGowan, C. *Nature* **307**, 733–735 (1984)
22 Howard, N. L., Joubert, F. J. & Strydom, D. J. *Comp. Biochem. Physiol.* **48B**, 75–85 (1974)
23 Oberthur, W., Braunitzer, G., Gnimm, F. & Kösters, J. *Hoppe-Seyler's Z. physiol. Chem.* **364**, 851–858 (1983)
24 Wyles, J. S., Kunkel, L. G. & Wilson, A. C. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4394–4397 (1983)

III-6: More information from αA-sequences

Despite the scarcity of amino acid replacements in most avian αA-sequences, we have attempted to extract some phylogenetic information from the group of "14 other avian species", left unresolved in the previous section.

Cladistic analysis, based upon 21 avian αA-sequences, groups these 14 species by three shared derived characters, i.e. the amino acid replacements 127 Ser → Ala, 135 Ser → Asn and 152 Pro → Ala, as can be seen on page 82.

Additional resolution may be obtained by considering 153 Ser → Gly to be a synapomorphy, clustering gannet, crow, eagle owl, buzzard and pigeon. Grouping of eagle owl and buzzard would be in accordance with the assumption of a monophyletic origin of strigiform and falconiform birds (see: Sibley and Ahlquist, 1972).

No synapomorphies can be attributed to the Charadriiformes, which group with coot, budgerigar, heron, fulmar and penquin. This indicates that another alleged close phylogenetic relationship, involving the Procellariiformes, here represented by fulmar, and Sphenisciformes (penguin) (Sibley and Ahlquist, 1972), is not contradicted by the constructed phylogenetic αA tree as presented in figure α-2, page 86.

It must be stressed, however, that different equally parsimonious topologies are possible as well on the basis of the observed αA amino acid replacements. Pigeon, as well as gull, for example, may be positioned with guillemot, because of 148 Ser → Ala, which would underline a previously assumed charadriiform-columbiform relationship (Sibley and Ahlquist, 1972). Buzzard can be placed with gull and curlew (32 Leu → Phe), but the latter two species might just as well be placed with buzzard.

In short, branching within the "14 other avian species"-cluster is not supported by conclusive evidence, obtained from αA-sequences, and additional macromolecular sequence data are needed to shed new light on these remaining obscured interordinal relationships.

As has been previously outlined, much controversy exists about the age of the avian assemblage (Chapter I). According to the "coelurosaurian ancestor-theory" the birds originated as a separate lineage about 140 million years ago, while the "pseudosuchian ancestor-theory" designates the avian stock to be about 220 million years old.

In this respect, comparative amino acid sequence determination might lend a helping hand again, because of the circumstance that macromolecular sequences possibly evolve at approximately constant rates (Zuckerkandl and Pauling, 1962). If this is the case, the number of sequence differences that have accumulated in each lineage since two species diverged from a common ancestor, divided by the time of divergence, gives the absolute rate of macromolecular evolution along each lineage. The divergence time has to be concluded in most cases from paleontological evidence. The rate of evolution may be considered as a "molecular clock", which can be used to calculate unknown times of species divergence.

According to these principles, the evolutionary rates of a number of polypeptide chains has been calculated, one of them being the αA-chain, which turns out to accumulate three amino acid replacements per 100 residues in 100 million years (De Jong et al., 1984).

When we consider the avian αA-chains in figure α-2, page 86, it is obvious that considerable variability in the rate of accumulation of replacements has occurred, a phenomenon which has also been observed in the study of mammalian αA-chains (De Jong, 1984). Some species (chicken, turkey, rhea) apparently have gathered relatively few substitutions (their αA-chains show 5 replacements), while others have undergone considerably more events (guillemot 13; penguin 12; gull, buzzard and pigeon: 10) since the avian-reptilian diver-

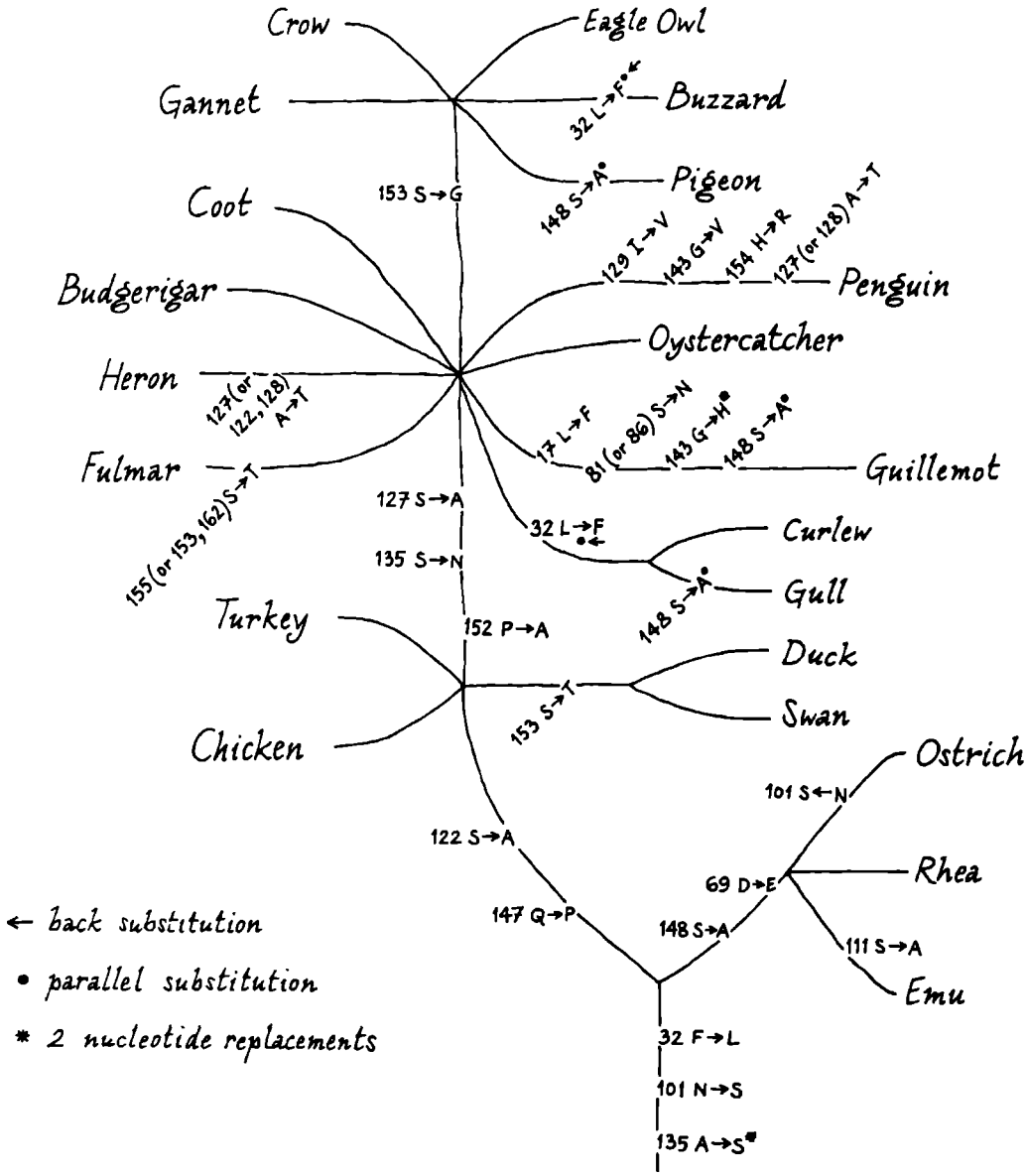


Figure α-2: Complete avian αA-tree based on the data set presented on page 82. Undetermined positions (...) are supposed to identical to the other avian residues, or filled in in such a way that they provide the least phylogenetic information.
This tree shows the least resolution possible. Many equally parsimonious alternatives are possible due to the frequent 0-branches. The constructed tree requires 29 amino acid substitutions (31 nucleotide replacements).

gence.

If one tries to calculate this avian-reptilian point of divergence, using the αA-evolution rate of 3 percent replacement in 100 million years, the average number of 4.7 amino acid replacements for avian αA-chains points to an avian-reptilian split-up of about 160 MYA. Since the variation of evolution rate in these avian chains is rather high ("the clock has run rather irregularly"), this calculated point of divergence is of doubtful value.

If one considers the traditionally accepted avian-reptilian divergence of about 200 MYA (Romer, 1966), avian αA-evolution seems to have been retarded in comparison to mammalian αA-evolution. This would be in line with an alleged general avian protein evolutionary slowdown (Prager and Wilson 1974; 1975), based upon paleontological and fossil evidence. This evidence, however, is being challenged (see: Wilson, 1977), and it can no longer definitely be excluded that modern birds evolved at a comparable rate as the mammalian stock, not only at the molecular, but also at the morphological level (Wyles et al., 1983). This possible equal evolution rate has also indirectly been suggested by Sibley and Ahlquist (1982) who commented, referring to their preliminary DNA-DNA hybridization results, on the inequivalence of the traditionally distinguished avian and mammalian taxonomic categories.

Whatever value these considerations may have, it appears to be possible that the avian assemblage is younger than has always been assumed, which is the most logical explanation for the mistaken "avian protein evolution slowdown" which may well be

> "... primarily the result of the limitations of human perception, not of some unknown difference between the genomes of birds and other animals" (Sibley and Ahlquist, 1982).

REFERENCES-III

- Bloemendal, H. et al. (1975) Exp. Eye Res. 20: 463-478.

- Cohen, L.H. et al. (1978) Eur. J. Biochem. 89: 259-266.

- De Jong, W.W. et al. (1981) Comp. Biochem. Physiol. 69B: 593-598.

- De Jong, W.W. et al. (1981) Nature 292: 538-540.

- De Jong, W.W. et al. (1984) Eur. J. Biochem. 141: 131-140.

- Ingolia, T.D. & Craig, E.A. (1982) Proc. Natl. Acad. Sci. USA 79: 2360-2364.

- King, C.E. & Piatigorsky, J. (1983) Cell 32: 707-712.

- Laemmli, U.K. (1970) Nature 227: 680-685.

- Prager, E.M. et al. (1974) J. Molec. Evol. 3: 243-262.

- Prager, E.M. & Wilson, A.C. (1975) Proc. Natl. Acad. Sci. USA 72: 200-204.

- Romer, A.S. (1966) "Vertebrate Paleontology", Univ. of Chicago Press, Chicago, 3th edition.

- Sibley, G.C. & Ahlquist, J. (1972) Bull. Peabody Mus. Nat. Hist. 39: 1-276.

- Sibley, G.C. & Ahlquist, J. (1982) "Postilla" 187: 1-19.

- Udenfriend, S. et al. (1972) Science 178: 871-872.

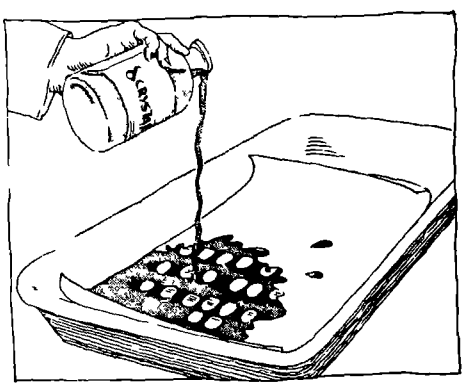
- Van Druten, H. et al. (1978) J. Theor. Biol. 73: 549-561.

- Wyles, J.S. et al. (1983) Proc. Natl. Acad. Sci. USA 80: 4394-4397.

- Zuckerkandl, E. & Pauling, L. (1962) in: "Horizons in Biochemistry" (Kasha, M. and Pullman, B. eds.) Academic Press, New York.

Chapter IV

δ-Crystallin



IV-1: Why δ-crystallin?

When the comparative avian αA amino acid sequence determinations were about to begin, it was realized that one particular circumstance could possibly jeopardize the whole project: the lack of sufficient accumulated amino acid replacements for the construction of a cladogram. This concern was not without reason: the αA-chain had been proven to be a relatively slowly evolving polypeptide (De Jong et al., 1977).

Bearing this in mind, a pilot study was undertaken to obtain some information about the evolutionary speed of another major water-soluble eye lens protein: δ-crystallin. This protein is, in contrast to α-crystallin, exclusively limited to the lenses of sauropsidan species (birds and reptiles).

This principal structural protein in the avian and reptilian (2) eye lens is a tetrameric aggregate in the 200 KDa molecular weight range and contains two types of polypeptide subunits, having molecular weights of 48 KDa and 50 KDa, respectively (Reszelbach et al., 1977).

The two different types of subunit were initially believed to be the products of the two δ-genes found in the genome, as was suggested in the case of chicken δ-protein (Bhat et al., 1980). Although these two genes, designated as δ1 and δ2, show outspoken resemblances, for example in their intron-exon patterns (Hawkins et al., 1984), recent evidence strongly indicates that only one gene, δ1, is being transcribed to a considerable degree. As was concluded from cDNA-sequencing, two initiation sites in the δ-gene sequence may be responsible for the existence of the two polypeptide chains of 48 KDa and 50 KDa (Nickerson et al., 1985).

If the two polypeptides are the products of the same gene in other sauropsidans as well, δ-crystallin might well be a suitable tool for the elucidation of avian phylogenetic relationships, using reptilian sequences for outgroup comparisons, provided that its evolutionary speed has been sufficiently high. For this reason, chicken, tegu and turtle δ-crystallin peptides were subjected to a comparative amino acid composition investigation to find out if this protein can be useful for evolutionary reconstructions (section IV-2).

Since the rate of evolution of δ-crystallin appeared to be markedly higher than the α-crystallin A evolution, while the protein can be obtained in high amounts from sauropsidan lenses, δ-crystallin must be a suitable tool for the elucidation of hitherto unresolved avian phylogenetic problems. This is especially facilitated by the recently reported δ-crystallin cDNA nucleotide sequences (Nickerson and Piatigorsky, 1984; Yasuda et al., 1984) from which an amino acid reference sequence can be deduced. It should therefore be possible to obtain some manageable CNBr-fragment(s) from different avian and reptilian species and to subject them to comparative amino acid sequence determination, in a similar way as has been described for the α-crystallin A chain.

δ-Crystallin evolutionary rate

Comp. Biochem. Physiol. 69B, 593-598 (1981)

Comp Biochem Physiol Vol 69B pp 593 to 598 1981
Printed in Great Britain

0305-0491 81 070593 06$02 00 0

A COMPARISON OF AVIAN AND REPTILIAN δ-CRYSTALLIN

WILFRIED W DE JONG, STEVEN O STAPEL and ANNEKE ZWEERS
Department of Biochemistry, University of Nijmegen, Geert Grooteplein Noord 21, 6525 EZ Nijmegen, The Netherlands

(*Received* 10 *November* 1980)

Abstract—1 The eye lens protein δ-crystallin is present in different amounts in different birds and reptiles It is absent in echidna

2 Chicken, tegu (*Tupinambis teguixin*) and turtle (*Chelonia mydas*) δ-crystallins have similar native and subunit mol wt Their amino acid compositions show the same characteristics

3 From peptide mapping and amino acid analyses the rate of evolutionary change of δ-crystallin is estimated to be 2–4 times faster than that of α-crystallin

INTRODUCTION

δ-Crystallin is a lens protein limited to birds and reptiles, initially designated as "first important soluble crystallin" (FISC) by Rabaey (1962) It has a mol wt between 150,000 and 200,000 and is composed of subunits with mol wt estimated between 45,000 and 50,000 (Rabaey *et al*, 1969, Piatigorsky *et al*, 1974, Thomson *et al*, 1978, Williams & Piatigorsky, 1979) The subunits demonstrate charge heterogeneity (Clayton, 1969, Thomson *et al*, 1978) and can also be resolved into two components in SDS-gel electrophoresis (Reszelbach *et al*, 1977) The different chick δ-crystallin subunits must be very similar in primary structure (Piatigorsky, 1976, Shinohara *et al*, 1980) and are encoded by at least two non-allelic genes (Bhat *et al*, 1980)

Very few comparative data are known about δ-crystallin, but interspecies differences of native δ-crystallin and its subunits have been found (Rabaey, 1962, Gysels, 1964, Clayton 1974) Also differences in primary and secondary structure do exist (Williams & Piatigorsky 1979, Horwitz & Piatigorsky, 1980) The main objective of the present study was to assess the degree of primary structure difference between avian and reptilian δ-crystallins This would allow an estimate of the rate of evolutionary change in comparison to the other crystallins

MATERIALS AND METHODS

Lenses were obtained from the following species chicken (*Gallus domesticus*) domestic pigeon (*Columba livia*) great northern diver (*Gavia immer*) common tegu (*Tupinambis teguixin*) cape monitor (*Varanus exanthematicus*) western diamond-back rattlesnake (*Crotalus atrox*), green turtle (*Chelonia mydas*) red-eared turtle (*Chrysemys scripta elegans*) and Australian echidna or spiny anteater (*Tachyglossus aculeatus*) All animals were adult or subadult Lenses of chicken pigeon, tegu monitor and red-eared turtle were obtained from the Central Animal Facilities University of Nijmegen School of Medicine, diver lenses from Dr J Wattel Department of Systematic Zoology University of Amsterdam snake lenses from Dr H M Verheij Department of Biochemistry University of Utrecht green turtle lenses from Dr J Wood Cayman Turtle Farm Cayman Islands echidna lenses from Dr P Zwart, Department of Veterinary Pathology University of Utrecht

Electrophoresis in the presence of sodium dodecylsulfate

Comparisons of total lens protein compositions of different species were made by slab-gel electrophoresis in 13% polyacrylamide gels containing 0 10% sodium dodecylsulfate (Laemmli, 1970)

Lenses were homogenized in 1% ammonium bicarbonate, pH 7 9 After centrifugation (10 at 15 000 *g*) an aliquot was taken from the supernatant and diluted 10-fold with a solution containing 2% SDS 5% β-mercaptoethanol, 10% glycerol, 0 005% bromophenol blue and a sample was applied on the gel The same electrophoretic technique was used to analyze fractionated lens proteins

Isolation of δ-crystallin

Lenses were homogenized in 2 vol of 1% ammoniumbicarbonate buffer, pH 7 9 After centrifugation (10 min at 15,000 *g*) the supernatant was applied to a column (120 × 3 5 cm) of Ultrogel AcA 34 (LKB) and eluted at room temperature with 1% ammoniumbicarbonate, pH 7 9, at a flow rate of 30 ml hr This procedure is comparable to the one described by Thomson *et al* (1978) apart from leaving out the β-mercaptoethanol from the buffer The fractions containing δ-crystallin were pooled and lyophilized Further purification was achieved by gel filtration under dissociating conditions for which the sample (50 mg) was suspended in 2 5 ml of 0 1 M sodium acetate buffer, 5 mM in dithiothreitol pH 8 5 saturated with urea, and dissolved by ultrasonic treatment The pH of the mixture was brought to 4 5 with 1 M acetic acid before applying it to a column (120 × 1 0 cm) of Sephacryl S-200 eluting it at room temperature at a flow rate of 6 ml h with a buffer containing 6 M urea 0 1 M sodium acetate and 2 mM dithiothreitol pH 4 5

Analysis of δ-crystallin

δ-Crystallin was reduced and aminoethylated (Raftery & Cole 1966) and digested with trypsin (Worthington TRTPCK) for 2 hr at 37 C The protein concentration was 10 mg ml of buffer (0 1 M ammoniumbicarbonate brought to pH 8 9 with ammonia) 1% (enzyme substrate w w) of trypsin was added at zero time and a similar portion after 1 hr The digestion was stopped by lowering the pH to 5 with 1 M HCl the insoluble core was removed by centrifugation and the supernatant lyophilized

Peptide maps were prepared from 2 5 mg samples of the digest using Whatman 3MM paper High voltage electrophoresis in pyridine acetic acid water buffer (25 1 225,

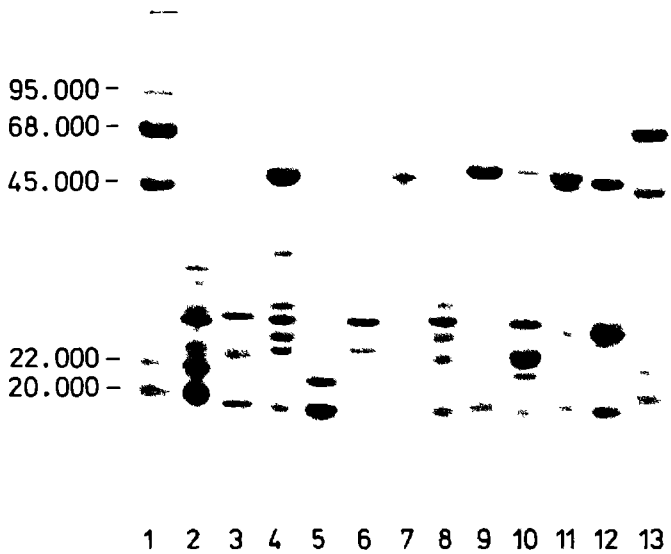


Fig 1 SDS-gel electrophoresis of marker proteins (lanes 1 and 13), and lens extracts of calf (2) echidna (3), chicken (4), pigeon (8) diver (9), red-eared turtle (10), monitor (11) and rattlesnake (12) Lanes 5 6 and 7 contain chicken α-, β- and δ crystallin, respectively Diver and pigeon are choosen from many screened avian species because of their very high and low levels of δ-crystallin, respectively

pH 6 5) was carried out for 75 min at 50 V cm, followed by descending chromatography in *n*-butanol/acetic acid/water/pyridine (15 3 12 10, v/v) for 18 hr The peptide maps were stained with 0 01% ninhydrin (w v) in acetone, containing 1% acetic acid and 1% pyridine or with fluorescamine (Fluram Roche) (Udenfriend *et al*, 1972) when peptides had to be isolated for amino acid analysis

Peptides were eluted from the paper with ½ ml of 6 M HCl, containing 0 025% (w v) phenol and hydrolyzed under vacuum for 22 hr at 110 C Amino acid analyses were performed on a Rank Hilger Chromaspek amino acid analyzer Samples of δ-crystallin were hydrolyzed for 24 48 and 72 hr and analyzed in duplo in the same way

RESULTS AND DISCUSSION

Distribution in different species

By virtue of its subunit mol wt of about 50,000 δ-crystallin can easily be distinguished from the other crystallins by SDS-gel electrophoresis (Fig 1) δ-Crystallin is present in greatly varying quantities in the total lens extracts of the investigated bird and reptile species Minor differences in mobility of the δ-crystallin subunits in different species can be seen, as well as the presence of multiple δ-bands in several species Similar observations have been made in other reptiles (python, gekko, caiman) and birds (quail, turkey, duck) by Williams & Piatigorsky (1979)

It is of interest that δ-crystallin appears to be absent in the lens of the echidna (Fig 1), because the monotremes are an isolated and primitive group of mammals which have many reptilian characters (Romer, 1966),

It has been suggested that δ-crystallin, which is unique among crystallins by being chiefly in the α-helical configuration, is responsible for the soft consistency and easy deformability of the sauropsidan lens (Yu *et al*, 1977) The presence of a large amount of δ-crystallin in the lens of the snake, which accommodates by displacement and not by deformation of the lens (Prince, 1956), lends no support to this suggestion Also the very high level of δ-crystallin in the diver and the very low level in the pigeon cannot simply be correlated with differences in accommodative capacity

Gel filtration of chicken and tegu crystallins

The water-soluble lens proteins of chicken and tegu were fractionated by gel filtration (Fig 2) and analyzed by SDS-gel electrophoresis (Fig 3) The similar elution volumes of native chicken and tegu δ-crystallin indicates that tegu δ-crystallin, like the avian (Rabaey *et al*, 1969) and caiman protein (Williams & Piatigorsky, 1979) has a native mol wt like that of chicken δ-crystallin

Both chicken and tegu show small amounts of high mol wt crystallin, mainly consisting of β-crystallin polypeptides in the chicken and α-crystallin in the tegu The α-crystallins of both species have slightly lower apparent subunit mol wt, and comparable ratios of αA to αB chains as in the calf (Fig 3)

The subunit compositions of chicken and tegu β-crystallins are quite different, although several bands have similar electrophoretic mobilities Our results are in general agreement with previous data on chicken and turtle β-crystallin (Zigler & Sidbury, 1976a Thomson *et al*, 1978) The last peak of the tegu elution pattern contains low mol wt protein which is not detectable in chicken, having an apparent mol wt of 23,000 This fraction may well correspond to the disputed "γ-crystallin" of the birds (McDevitt & Croft, 1977) Zigler & Sidbury (1976a)

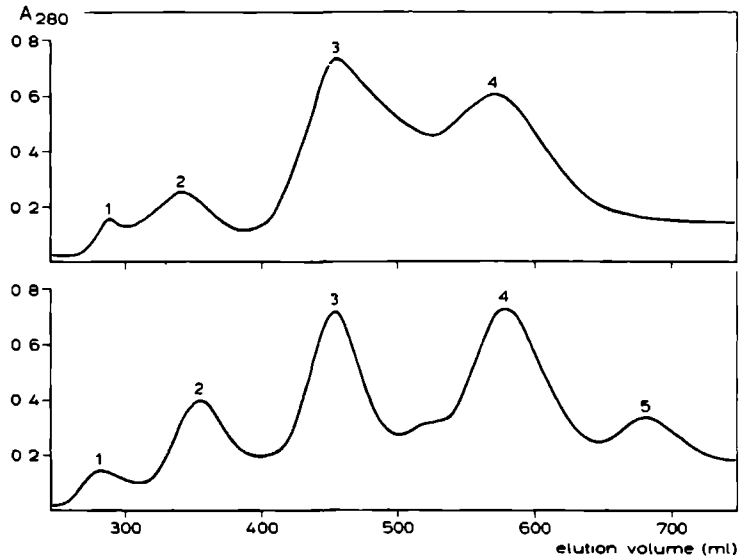


Fig 2 Fractionation by gel filtration of water-soluble lens proteins of chicken (top) and tegu (bottom) The indicated peak fractions were analyzed by SDS-gel electrophoresis (Fig 3)

observed a single 20,000 mol wt component in the snapping turtle low mol wt fraction

Analysis of δ-crystallins

After gel filtration the δ-crystallin fractions of both chicken and tegu still contained considerable amount, of β_{High}-crystallin (Fig 3) Taking advantage of the difference in monomer molecular weights of δ- and β-crystallin, the δ-crystallin subunits could be further purified by gel filtration under dissociating conditions, yielding essentially pure δ-crystallin (Fig 3)

The amino acid compositions of δ-crystallins from chicken, tegu and green turtle (isolated in small quantity by the same procedures) show clear similarities

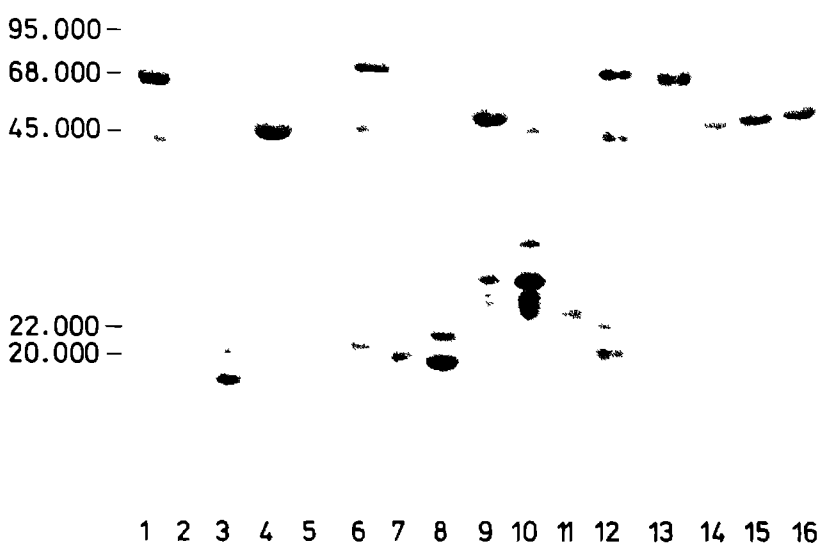


Fig 3 SDS-gel electrophoresis of the peak fractions indicated in Fig 2, and of purified δ-crystallins Marker proteins (lanes 1, 6 12 and 13), chicken high-mol wt crystallin (lane 2), α-crystallin (3), δ-plus β_H-crystallin (4), β_L-crystallin (5), tegu high-mol wt crystallin (7), α-crystallin (8), δ- plus β_H-crystallin (9), β_L-crystallin (10), "′"-crystallin (11), purified δ-crystallin from chicken (14), green turtle (15) and tegu (16)

Table 1 Amino acid compositions of chicken and reptilian δ-crystallins (residues 1000 residues)

	Chicken	Tegu	Turtle
Asp	77 (71)*	67	95
Thr	72 (76)	70	54
Ser	88 (96)	97	83
Glu	130 (130)	151	110
Pro	15 (23)	21	15
Gly	61 (56)	76	80
Ala	88 (80)	84	83
Cys	nd (3)	nd	nd
Val	69 (80	72	69
Met	18 (8)	9	34
Ile	70 (74)	71	51
Leu	145 (151)	129	109
Tyr	13 (8)	17	24
Phe	26 (22)	16	43
His	13 (12)	19	27
Lys	83 (72)	92	71
Arg	56 (39)	48	75

Values are the average of duplicate analyses after 24 48 and 72 hr of hydrolysis Values for threonine and serine are obtained by extrapolation to zero time hydrolysis valine and isoleucine have values for 72 hr of hydrolysis

* Values obtained by Piatigorski *et al* (1974)

(Table 1) The analysis of chicken δ-crystallin is in good agreement with the values reported by Piatigorsky *et al* (1974) Characteristic for all δ-crystallins are the low values for proline and the high values for leucine and lysine These specific compositional features as compared to the other crystallins and to the "average" composition of proteins are obvious from Fig 4 They certainly can be attributed in part to the particular conformational properties of δ-crystallin

The fingerprints of the soluble tryptic peptides of the aminoethylated δ-chains of chicken and tegu (Fig 5) show fewer peptides than expected from a chain of approximately 420 residues containing some 35 lysyl and 24 arginyl residues Large parts of the chain must be present in the insoluble core remaining after tryptic digestion Some similarity can be seen between the peptide maps of chicken and tegu δ-crystallin All peptides were analyzed but many of them were too large or contaminated to be useful for reliable comparisons between the two species Only 15 mostly small peptides (indicated in Fig 5) were found to be clearly homologous in the two species and their compositions are aligned in Table 2 At least 13 amino acid substitutions are present among the 71 residues represented by these peptides This means, for these parts of the δ-chain a degree of sequence difference of at least 18% These peptides however being small and with easily detectable homology are unlikely to be representative for the total chain of 420 residues The over-all sequence difference may well be higher than 18%

Rate of evolutionary change of δ-crystallin

Two other approaches may help to estimate the rate of sequence change of δ-crystallin Cornish-Bowden (1979) has shown that the sequence difference between two homologous proteins can be estimated from the differences in their amino acid compositions Applying her formula to the compositional difference between chicken and tegu δ-crystallin indicates a 32% sequence difference

Also the amino acid composition itself of a protein shows some correlation with the rate of evolution of that protein (Chirpich 1975) On this basis the relative mutability of δ-crystallin should be higher than

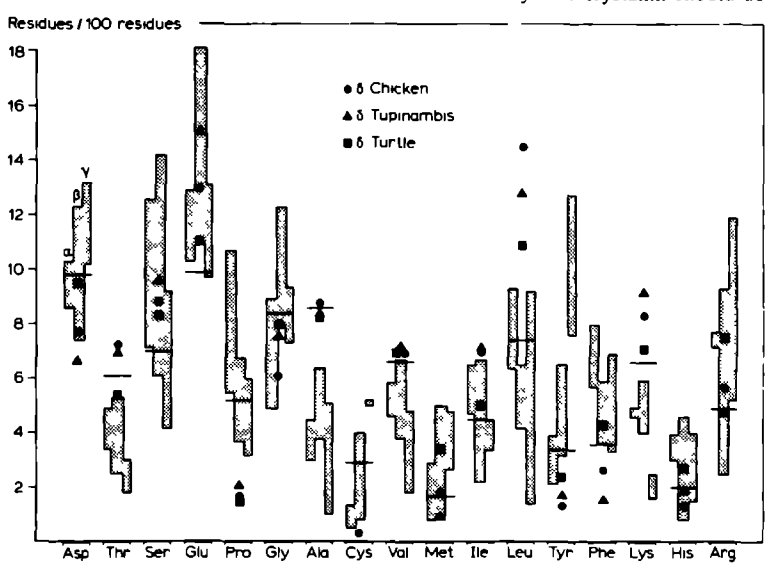


Fig 4 Amino acid compositions of chicken tegu and green turtle δ-crystallins (from Table 1) compared to the range of reported amino acid compositions of vertebrate α- β- and γ-crystallins (shaded areas) and the average % of amino acids in proteins (horizontal bars) (Dayhoff 1978 p 363) Included are the compositions of calf human chicken and dogfish α-crystallins, calf sheep pig rat rabbit rhesus monkey, chicken turtle frog bluefish and dogfish β-crystallins bovine β_S-crystallin calf human rat haddock and dogfish γ-crystallins pigeon γ -crystallin (included in the β-crystallins) (McDevitt & Croft 1977, Zigler & Sidbury 1976 a,b and references in De Jong, 1981)

that of the crystallin chains αA, αB, βBp and δ-Crystallin would thus be expected to be the fastest evolving among the evolutionary generally conservative crystallins (De Jong, 1981)

The values of 18 32% sequence difference between chicken and tegu δ-crystallin are indeed 2–4 times higher than the 8 5% sequence difference established between the α-crystallin A chains of chicken and tegu (De Jong & Zweers, unpublished data) The rate of evolutionary change of δ-crystallin might then be esti-

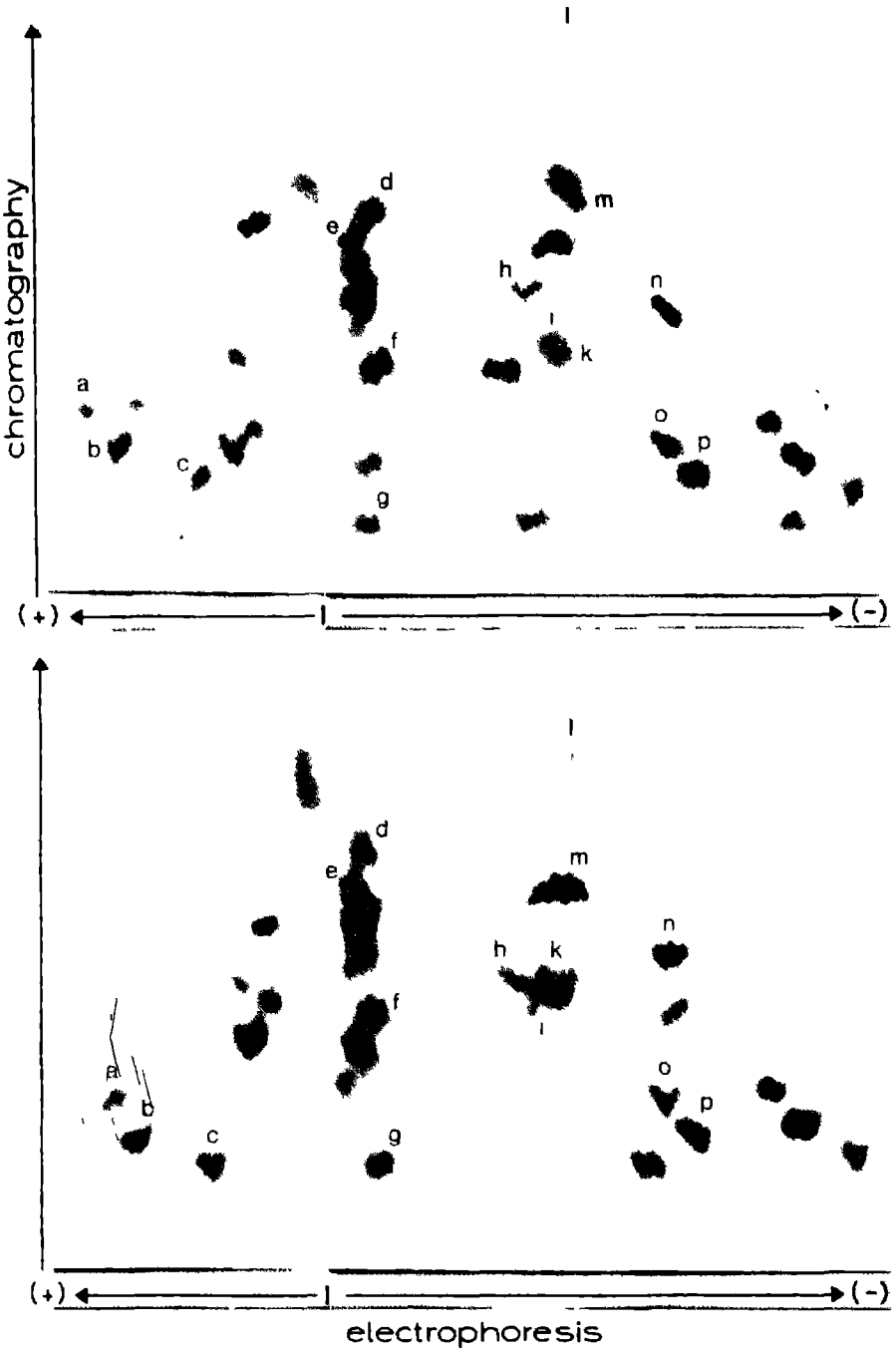


Fig 5 Fingerprints of the soluble tryptic peptides of chicken (top) and tegu (bottom) aminoethylated δ-crystallin Homologous peptides of which the compositions are given in Table 2, are indicated by the letters

Table 2 Compositions of supposedly homologous tryptic peptides from chicken and tegu δ-crystallins (as indicated in Fig 5)

Peptide	a	b	c
Chicken	Asp Thr Glu Gly Ala Lys	Asx Asx Glx Glx Leu Lys	Asp Ser Glu Arg
Tegu	Glu Ser – — — – —		– — — —

Peptide	d	e	f	g
Chicken	Ser Glu Gly Ile Leu Leu Arg	Glu Leu Leu Arg	Glu Ala Leu Lys	Glx Glx Ala Lys
Tegu	Gly —		— — — –	– -- Gly

Peptide	h	i	k
Chicken	Thr Ser Ser Pro Gly Ile Phe Lys	Glu Leu Leu Lys Lys	Thr Gln Gly Leu Arg
Tegu	Asn Leu Tyr — – Ile –		Pro Ala Ile

Peptide	l	m	n	o	p
Chicken	Ala Ala Leu Leu Leu Lys Lys	Gly Val Phe Arg	Leu Lys	Gly Ala Arg	Ser Lys
Tegu	Met — – —	Met			– —

Peptides are aligned for maximum homology on the basis of amino acid compositions Residues are arranged in the order of elution from the column Amides or carboxyl groups are assigned where possible on the basis of electrophoretic mobility of the peptides

mated between 6 and 12 substitutions per 100 residues in 100 Myr, being 2–4 times faster than the rate of 3% sequence change per 100 Myr for αA (De Jong *et al*, 1980)

Acknowledgements Drs J Wattel (Amsterdam), J Wood (Cayman Islands), H M Verheij and P Zwart (Utrecht) kindly supplied eye lenses We are grateful to Marlies Versteeg for performing the amino acid analyses The investigations were supported in part by the Foundation for Fundamental Biological Research (BION) and the Netherlands Foundation for Chemical Research (SON), which are subsidized by the Netherlands Organization for Pure Research (ZWO)

REFERENCES

BHAT S P, JONES R E SULLIVAN M A & PIATIGORSKY J (1980) Chicken lens crystallin DNA sequences show at least two δ-crystallin genes *Nature* **284**, 234–238

CHIRPICH T P (1975) Rates of protein evolution a function of amino acid composition *Science* **188**, 1022–1023

CLAYTON R M (1969) Properties of the crystallins of the chick in terms of their subunit composition *Expl Eye Res* **8**, 326–339

CLAYTON R M (1974) Comparative aspects of lens proteins In *The Eye* (Edited by DAVSON H & GRAHAM I T) Vol 5 pp 399–494 Academic Press London

CORNISH-BOWDEN A (1979) How reliable do amino acid composition comparisons predict sequence similarities between proteins? *J Theor Biol* **76**, 369–386

DAYHOFF M O (1978) *Atlas of Protein Sequence and Structure* Vol 5, Suppl 3 National Biomedical Research Found, Silver Spring, Maryland

DE JONG W W (1981) Evolution of lens and crystallins In *Molecular and Cellular Biology of the Eye Lens* (Edited by BLOEMENDAL H) Wiley-Interscience, New York pp 221–278

DE JONG W W, ZWEERS A & GOODMAN M (1980) Trends in the molecular evolution of α-crystallin In *Protides of the Biological Fluids* (Edited by PETERS H) Vol 28 pp 161–164 Pergamon Press, Oxford

GYSELS H (1964) Immunoelectrophoresis of avian lens proteins *Experientia* **20**, 145–146

HORWITZ J & PIATIGORSKY J (1980) Evolutionary and developmental differences in δ-crystallin from bird and reptile lenses Circular dichroism and fluorescence studies *Biochim biophys Acta* **624**, 21–29

LAEMMLI U K (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4 *Nature* **227**, 680–685

MCDEVITT D S & CROFT L R (1977) On the existence of γ-crystallin in the bird lens *Expl Eye Res* **25**, 473–481

PIATIGORSKY J (1976) Subunit composition of δ-crystallin from embryonic chick lens *J biol Chem* **251**, 4416–4420

PIATIGORSKY J, ZELENKA P & SIMPSON R T (1974) Molecular weight and subunit structure of δ-crystallin from embryonic chick lens fibers *Expl Eye Res* **18**, 435–446

PRINCE J H (1956) *Comparative Anatomy of the Eye*, pp 228–242 Thomas, Springfield Il

RABAEY M (1962) Electrophoretic and immunoelectrophoretic studies on the soluble proteins in the developing lens of birds *Expl Eye Res* **1**, 310–316

RABAEY M, LAGASSE A & DE METS M (1969) Comparative studies of the lens proteins of birds *Acta zool pathol Antverp* **48**, 63–71

RAFTERY M A & COLE R D (1966) On the aminoethylation of proteins *J biol Chem* **241**, 3457–3461

RESZELBACH R SHINOHARA T & PIATIGORSKY J (1977) Resolution of two distinct embryonic chick δ-crystallin bands by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate and urea *Expl Eye Res* **25**, 585–593

ROMER A S (1966) *Vertebrate Paleontology* 3rd edn pp 197–198 Univ of Chicago Press MI

SHINOHARA T RESZELBACH R & PIATIGORSKY J (1980) Two tryptic peptide differences among the subunits of δ-crystallin of the embryonic chick lens *Expl Eye Res* **30**, 361–370

THOMSON R WILKINSON C E BURNS A T H TRUMAN D E S & CLAYTON R M (1978) Characterization of chick lens soluble proteins and the control of their synthesis *Expl Eye Res* **26**, 351–362

UDENFRIEND S STEIN S BOHLEN P DAIRMAN W LEIMGRUBER W & WEIGELE M (1972) Fluorescamine a reagent for assay of amino acids in the picomole range *Science* **178**, 871–872

WILLIAMS L A & PIATIGORSKY J (1979) Comparative and evolutionary aspects of δ-crystallin in the vertebrate lens *Eur J Biochem* **100**, 349–357

YU N-T, FAST E J, CHANG R C C & KUCK J F R (1977) Raman spectra of bird and reptile lens proteins *Expl Eye Res* **24**, 321–334

ZIGLER J S & SIDBURY J B (1976a) A comparative study of the β-crystallins of four sub-mammalian species *Comp Biochem Physiol* **55B**, 19–24

ZIGLER J S & SIDBURY J B (1976b) A comparative study of β-crystallin from six mammals *Comp Biochem Physiol* **53B**, 349–355

REFERENCES-IV

- Bhat, S.P. et al. (1980) Nature 284: 234-238.

- De Jong, W.W. et al. (1977) J. Molec. Evol. 10: 123-135.

- Hawkins, J.W. et al. (1984) J. Biol. Chem. 259: 9821-9825.

- Nickerson, J.M. & Piatigorsky, J. (1984) Proc. Natl. Acad. Sci. USA 76: 3299-3303.

- Nickerson, J.M. et al. (1984) J. Biol. Chem 260: 9100-9105.

- Reszelbach, R. et al. (1977) Exp. Eye Res. 25: 583-593.

- Williams, L.A. et al. (1982) Biochem. Biophys. Acta 708: 49-56.

- Yasuda, K. et al. (1984) EMBO J. 3: 1397-1402.

CONCLUDING REMARKS

Cladistic analysis of 21 avian αA-chains has resulted in taxonomically important conclusions. Nevertheless, this polypeptide has not accumulated sufficient amino acid replacements to provide satisfactory answers to many unresolved avian phylogenetic problems.

Since comparative studies of orthologous macromolecular sequences can offer suitable characters for taxonomic analyses, the avian phylogenetic questions may well be solved by these molecular approaches.

I therefore recommend two possibilities:

1. Comparative amino acid sequence determination of δ-crystallin, for reasons outlined on page 93.

2. Comparative nucleotide sequence determination of orthologous genomic DNA-fragments; for instance restriction fragments comprising the single-copy gene for α-crystallin A.
   Such fragments contain, apart from the nucleotide sequence coding for the amino acid chain, much more sequential information which is not reflected in the polypeptide, such as intervening sequences and possible "third base"-substitutions.
   The DNA-sequence, in other words, can provide an almost infinite number of characters, suitable for taxonomic and phylogenetic analyses.

CURRICULUM VITAE

Steven Stapel werd op 9 maart 1949 geboren te Amsterdam, en doorliep de Gymnasium-β opleiding aan het Gemeentelijk Lyceum te Eindhoven. In 1969 werd de studie Biologie begonnen aan de Universiteit van Amsterdam. Het kandidaatsexamen B_1 werd behaald in juni 1974, waarna in mei 1979 het doctoraalexamen werd afgelegd met als hoofdvak Plantenfysiologie (Prof.Dr. D. Stegwee), als bijvak Aquatische Oecologie (Prof.Dr. J. Ringelberg) en de neverichtingen Biochemie en Onderwijskunde.

Van 1 juni 1979 tot 1 juli 1982 was auteur als wetenschappelijk medewerker in dienst van de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek (ZWO) verbonden aan het Laboratorium voor Biochemie (Prof.Dr. H. Bloemendal) van de Katholieke Universiteit te Nijmegen, om het onderhavig promotieonderzoek te verrichten onder leiding van Dr. W.W. de Jong en Dr. J. Wattel (Instituut voor Taxonomische Zoölogie van de Universiteit van Amsterdam), in het kader van een project van de Stichting Biologisch Onderzoek Nederland (BION).

Na een korte periode onder leiding van Dr. W.J. van Venrooij gewerkt te hebben aan de karakterisering van autoantigenen bij rheumatische ziekten (aan het Laboratorium voor Biochemie van de Katholieke Universiteit te Nijmegen), werd auteur aangesteld als hoofd Allergie-diagnostiek van het Centraal Laboratorium van de Bloedtransfusiedienst te Amsterdam.

STELLINGEN

I

De conclusie van Lee et al., dat reconstructie van fylogenetische relaties en evolutionaire gebeurtenissen aan de hand van de primaire structuur van slechts een enkel eiwit of gen onbetrouwbaar is, wordt niet ondersteund door hun eigen bevindingen, daar deze berusten op een alignment-fout in de gebruikte eiwitdataset.

Lee, Y.M., Friedman, D.J. and Ayala, F.J. (1985) Proc. Natl. Acad. Sci. USA 82; 824-828.

II

Constructie van evolutionaire bomen op basis van hydrofobiciteitswaarden der betrokken aminozuren kan in voorkomende gevallen van doorslaggevende betekenis zijn.

Leunissen, J.A.M. and De Jong, W.W. (1986) J. Theor. Biol. (in press).

III

Het verdient aanbeveling om bij rheumatische aandoeningen meer gedetailleerd onderzoek te doen naar de correlatie van klinische en serologische parameters.

IV

De door Owen et al. beschreven Pittsburgh-variant van α_1-antitrypsine illustreert dat bij patienten met recidiverende bloedingen, met name indien deze optreden tijdens acuut-fase reacties, onderzoek gedaan moet worden naar de aanwezigheid van gemuteerde proteaseremmers.

Owen, M.C., Brennan, S.O., Lewis, J.H. and Carrell, R.W. (1983) N. Engel. J. Med. 309; 694-698.

V

Het is geenszins bewezen dat de door Agostini et al. beschreven monoclonale antistof daadwerkelijk met kallikreine-C1-esteraseremmercomplexen reageert, aangezien geen rekening gehouden is met mogelijke dissociatie van deze complexen.

Agostini, A., Schapira, M., Wachtfogel, Y.T., Colman, R.W. and Carrel, S. (1985) Proc. Natl. Acad. Sci. USA 82; 5190-5193.

VI

De discrepantie tussen immunochemisch en hemolytisch gekwantiteerd C1q in het supernatant van in vitro rijpende monocyten berust op de secretie van Laag-Moleculair-C1q door deze cellen.

Tenner and Volin (1986) Biochem. J. 233; 451-458
Hoekzema R. et al. (1986) Annual Report 1985 of the Dr. Karl Landsteiner Foundation (in press).

VII

Het gebruik van een anti-hapteen antistof, gekoppeld aan een vaste drager, in combinatie met een hapteen-gekoppeld antigeen is een in brede zin toepasbaar alternatief voor de momenteel toegepaste antistofbepalingen die berusten op een directe koppeling van antigeen aan vaste drager.

Aalberse, R.C., Van Zoonen, M., Clemens, J.G.J. and Winkel, I. (1986) J. Immunol. Methods (in press).

VIII

Bij de presentatie van de resultaten van het evaluatie-onderzoek naar de betrouwbaarheid van commercieel verkrijgbare immuno assays voor het aantonen van antistoffen tegen LAV/HTLV-III heeft men ten onrechte de onderling sterk uiteenlopende gevoeligheid van deze testen buiten het eindoordeel gehouden.

CMBC-symposium "Donorscreening op anti-LAV/HTLV-III", 's-Gravenhage (1985).

IX

Onderzoek naar antigeen-neutraliserende antistoffen tegen het AIDS virus zijn alleen zinvol in experimentele diermodellen.

X

Het begrip "allergie" dient beperkt te blijven tot overgevoeligheden op immunologische basis.

XI

Inhalatie van graspollen kan allergie voor aardappelen veroorzaken.

XII

Ook na dit proefschrift zullen de controverses omtrent de fylogenie der ratites voortduren.

XIII

Nieuw-Guinea had niet aan Indonesië overgedragen mogen worden.

"Birds and Biochemistry, it's getting time for them to meet again"

S. Stapel, '86